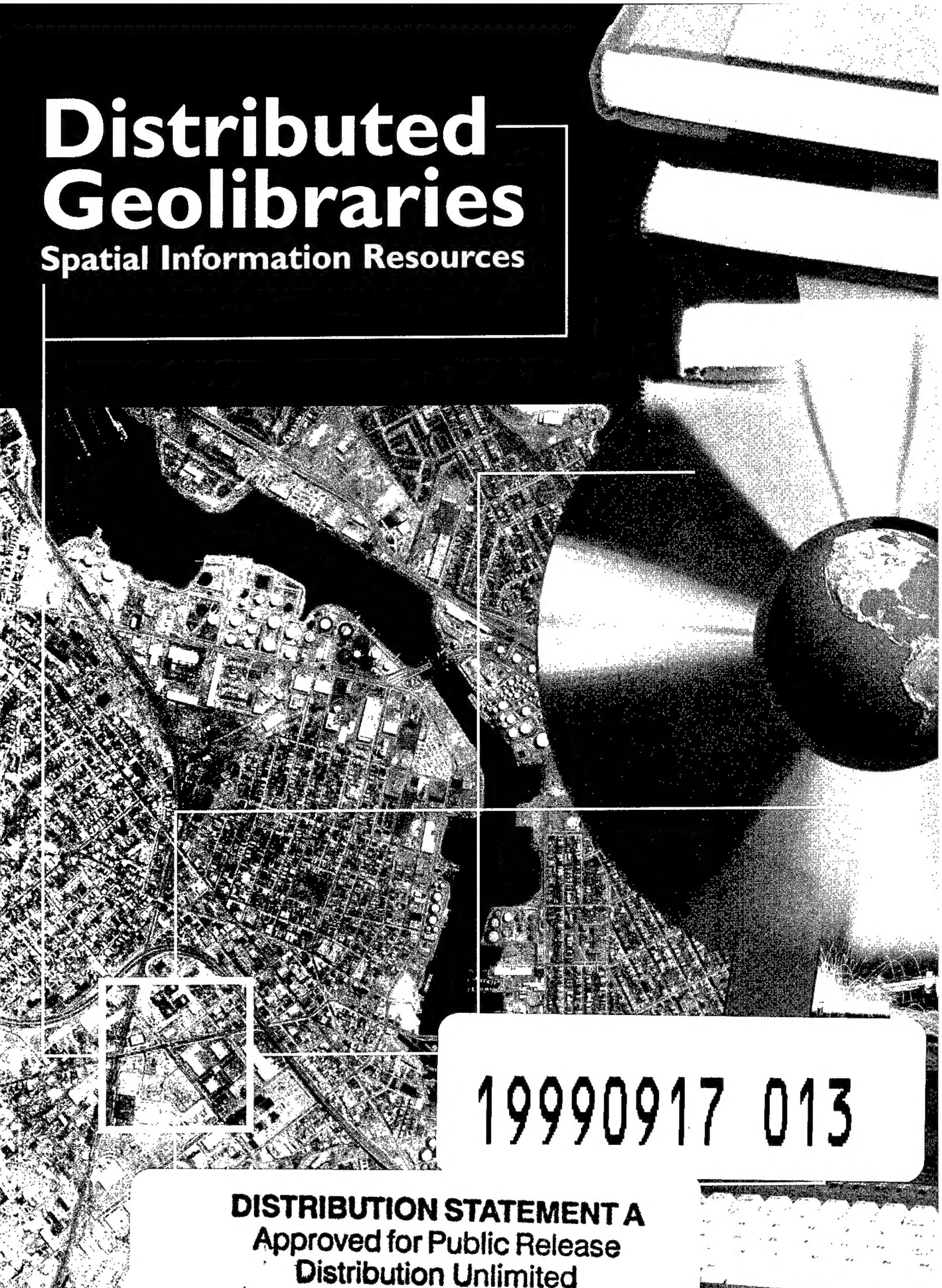
Distributed Geolibraries

Spatial Information Resources



NATIONAL RESEARCH COUNCIL

Distributed Geolibraries

Spatial Information Resources

Summary of a Workshop

Panel on Distributed Geolibraries

Mapping Science Committee

Board on Earth Sciences and Resources

Commission on Geosciences, Environment, and Resources

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.

Support specifically for this project was provided by the National Science Foundation and the Defense Advanced Research Projects Agency. The project also utilized resources provided to the Mapping Science Committee by the National Imagery and Mapping Agency, the U.S. Geological Survey and the Federal Geographic Data Committee, the Bureau of Transportation Statistics, the National Oceanic and Atmospheric Administration, the Bureau of Land Management, and the Bureau of the Census. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the agencies that provided support for this project.

International Standard Book Number (ISBN) 0-309-06540-2

Copies of this report are available from

Mapping Science Committee
Board on Earth Sciences and Resources
National Research Council
2101 Constitution Avenue, NW
Washington, DC 20418

Cover: Backdrop for the collage is a digital orthophoto of the Boston, Massachusetts, area. The figure was downloaded from the Internet from MIT/MassGIS Digital Orthophoto Project (see Appendix D)

Printed in the United States of America

PANEL ON DISTRIBUTED GEOLIBRARIES

MICHAEL F. GOODCHILD (*Chair*) University of California, Santa Barbara
PRUDENCE S. ADLER, Association of Research Libraries, Washington, D.C.
BARBARA P. BUTTENFIELD, University of Colorado, Boulder
ROBERT E. KAHN, Corporation for National Research Initiatives, Reston, Virginia
ANNETTE J. KRYGIEL, National Defense University, Ft. Lesley J. McNair, Washington, D.C.
HARLAN J. ONSRUD, University of Maine, Orono

NRC Staff

THOMAS M. USSELMAN, Senior Staff Officer
JENNIFER T. ESTEP, Administrative Assistant

MAPPING SCIENCE COMMITTEE

MICHAEL F. GOODCHILD (*Chair*) University of California, Santa Barbara
KAREN C. SIDERELIS (*Vice-Chair*) North Carolina Center for Geographic Information and Analysis, Raleigh
BRIAN J. L. BERRY, The University of Texas at Dallas
CLIFFORD A. BEHRENS,[+] Telcordia Technologies, Morristown, New Jersey
BARBARA P. BUTTENFIELD,[*] University of Colorado, Boulder
NICHOLAS CHRISMAN, University of Washington, Seattle
DAVID J. COLEMAN, University of New Brunswick, Fredericton
MICHAEL J. FOLK,[*] University of Illinois, Urbana
HENRY L. GARIE, New Jersey Department of Environmental Protection, Trenton
BARRY GLICK, Carillon Consulting, Arlington, Virginia
NINA S-N. LAM, Louisiana State University, Baton Rouge
JOEL L. MORRISON,[+] Ohio State University, Columbus
HARLAN J. ONSRUD, University of Maine, Orono
C. STEPHEN SMYTH, Microsoft Corporation, Redmond, Washington
REX W. TRACY, GDE Systems, Inc., San Diego, California
A. KEITH TURNER, Colorado School of Mines, Golden
LYNA L. WIGGINS, Rutgers University, New Brunswick, New Jersey

NRC Staff

THOMAS M. USSELMAN, Senior Staff Officer
JENNIFER T. ESTEP, Administrative Assistant

[*]Term of appointment ended December 31, 1998.
[+]Term of appointment began in 1999.

BOARD ON EARTH SCIENCES AND RESOURCES

COMMISSION ON GEOSCIENCES, ENVIRONMENT, AND RESOURCES

GEORGE M. HORNBERGER (*Chair*) University of Virginia, Charlottesville
RICHARD A. CONWAY, Union Carbide Corporation (*Retired*), S. Charleston, West Virginia
THOMAS E. GRAEDEL, Yale University, New Haven, Connecticut
THOMAS J. GRAFF, Environmental Defense Fund, Oakland, California
EUGENIA KALNAY, University of Oklahoma, Norman
DEBRA KNOPMAN, Progressive Policy Institute, Washington, D.C.
KAI N. LEE, Williams College, Williamstown, Massachusetts
RICHARD A. MESERVE, Covington & Burling, Washington, D.C.
JOHN B. MOONEY, JR., J. Brad Mooney Associates, Ltd., Arlington, Virginia
HUGH C. MORRIS, El Dorado Gold Corporation, Vancouver, British Columbia
H. RONALD PULLIAM, University of Georgia, Athens
MILTON RUSSELL, University of Tennessee, Knoxville
THOMAS C. SCHELLING, University of Maryland, College Park
ANDREW R. SOLOW, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts
VICTORIA J. TSCHINKEL, Landers and Parsons, Tallahassee, Florida
E-AN ZEN, University of Maryland, College Park
MARY LOU ZOBACK, U.S. Geological Survey, Menlo Park, California

NRC Staff

ROBERT M. HAMILTON, Executive Director
GREGORY H. SYMMES, Associate Executive Director
CRAIG SCHIFFRIES, Associate Executive Director for Special Programs
JEANETTE SPOON, Administrative and Financial Officer
SANDI FITZPATRICK, Administrative Associate
MARQUITA SMITH, Administrative Assistant/Technology Analyst

ACKNOWLEDGMENT OF REVIEWERS

This report has been reviewed by individuals chosen for their diverse perspectives and technical expertise in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the authors and the NRC in making their published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The content of the review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their participation in the review of this report:

Christine L. Borgman
Presidential Chair in Information Studies
University of California, Los Angeles

Edward A. Fox
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg

Kenneth D. Gardels
Research Program in Environmental Planning
and Geographic Information Systems
College of Environmental Design
University of California, Berkeley

John L. King
Department of Information and Computer Science
University of California, Irvine

Xavier R. Lopez
Spatial Products/Data Server Division
Oracle Corporation
Nashua, New Hampshire

Clifford A. Lynch
Executive Director
Coalition for Networked Information
Washington, D.C.

Hugh C. Morris
El Dorado Gold Corporation
Vancouver, British Columbia

Jane Smith Patterson
Senior Advisor for Science and Technology
Office of the Governor
Raleigh, North Carolina

James F. Williams II
Dean of Libraries
University of Colorado, Boulder

While the individuals listed above have provided many constructive comments and suggestions, responsibility for the final content of this report rests solely with the authoring committee and the NRC.

PREFACE

The Mapping Science Committee serves as a focus for external advice to federal agencies on scientific and technical matters related to spatial data handling and analysis. The purpose of the committee is to provide advice on the development of a robust national spatial data infrastructure for making informed decisions at all levels of government and throughout society in general.

The concept of a national spatial data infrastructure (NSDI) was first advanced by the Mapping Science Committee (MSC) in its 1993 report, *Toward a Coordinated Spatial Data Infrastructure for the Nation*. Subsequent MSC reports have addressed specific components of the NSDI, including partnerships (*Promoting the National Spatial Data Infrastructure Through Partnerships*, 1994), basic data types (*A Data Foundation for the National Spatial Data Infrastructure*, 1995), and future trends (*The Future of Spatial Data and Society*, 1997).

When the NSDI was defined in 1993, few users or producers of geospatial data* made much use of the Internet or the World-Wide Web (WWW). Although there was emphasis on digital geospatial data, the primary method of dissemination was by magnetic tape. There were virtually no digital online catalogs of geospatial data or methods for searching for data across computer networks. Moreover, since most useful geospatial data were produced by a small number of federal agencies, there was little problem locating the appropriate source. Today, the WWW has grown into an enormously successful tool and has had a profound impact on the entire environment for geospatial data acquisition. At the same time, it has presented a growing problem as the number of potential suppliers has mushroomed, in its inability to deal effectively with the task of

* The report follows evolving practice in the NSDI community by adopting the term *geospatial* to refer to maps and images of the Earth's surface and near surface and their digital equivalents. The terms *geographic* and *spatial* are often used almost synonymously but are avoided here.

discovering what geoinformation exists and of locating an appropriate source.

This report can be understood therefore as an updating of the MSC's concept of the NSDI in the era of the WWW. In organizing this effort and producing this report, the committee is expressing its view that the WWW has added a new and radically different dimension to its earlier conception of NSDI, one that is much more user oriented, much more effective in maximizing the value of the nation's geospatial data assets, and much more cost effective as a data dissemination mechanism. Distributed geolibraries reflect the same basic thinking about the future of geospatial data, which emphasizes sharing, universal access, and productivity but in the context of a technology that was almost impossible to anticipate prior to 1993.

A panel under the aegis of the MSC convened a workshop to explore the following topics:

- Development of a vision for geospatial data dissemination and access in 2010.
- Comparison of current efforts in digital library research, clearinghouse development, and other data distribution and search activities.
- Suggestion of short- and long-term research and development needed to achieve the vision.
- Identification of the policy and institutional issues, particularly for convergence of efforts to realize the vision.

By clarifying the vision of distributed geolibraries and identifying some of the key issues, it is hoped that the workshop and this report will provide a common focus for the many efforts already under way and will stimulate new and expanded efforts. The workshop was only a first step in this process, and many issues remain to be clarified by further discussions, research, and development of prototypes.

The report makes extensive use of the traditional library as a framework for discussion because it is so familiar and well understood. Undoubtedly, much future work in researching and developing distributed geolibraries will occur within this framework,

but the framework will also be constraining in some respects. Exactly how distributed geolibraries develop and how closely they follow the metaphor of the library remain to be seen. Moreover, the metaphor is used selectively, since many of the functions of libraries that may have no equivalent in distributed geolibraries were not discussed at the workshop, and may not be relevant.

The workshop began on Monday, June 15, 1998, and followed the agenda given in Appendix C. Workshop participants were selected in such a way that all major sectors of the NSDI community and geospatial data activity were represented by their respective stakeholders, with an appropriate balance among them. Of the participants, 35 percent were from federal and state government, 39 percent were from academia, 12 percent were from the private sector, and 14 percent were from other sectors (e.g., associations). See Appendix A for a list of participants. Another way of considering the participants is by their primary focus—44 percent with a geospatial background, 36 percent from computing science and engineering, 12 percent from the library sciences, and 8 percent "other."

The Panel on Distributed Geolibraries coordinated the preparation of a series of white papers in advance of the workshop to stimulate discussion on certain key issues. These were posted on the WWW several weeks prior to the workshop and were available to participants and others who happened across them. Titles of the white papers for the workshop are given in Appendix B.

This report reflects the consensus of the panel regarding the discussions that took place at the workshop, the issues that arose there and in the white papers, and the workshop's broader context.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is interim president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce Alberts and Dr. William A. Wulf are chairman and interim vice-chairman, respectively, of the National Research Council.

CONTENTS

Executive Summary

A distributed geolibrary is a vision for the future. It would permit users to quickly and easily obtain all existing information available about a place that is relevant to a defined need. It is modeled on the operations of a traditional library, updated to a digital networked world, and focused on something that has never been possible in the traditional library: the supply of information in response to a *geographically* defined need. It would integrate the resources of the Internet and the World Wide Web into a simple mechanism for searching and retrieving information relevant to a wide range of problems, including natural disasters, emergencies, community planning, and environmental quality. A geolibrary is a digital library filled with geoinformation—information associated with a distinct area or footprint on the Earth's surface—and for which the primary search mechanism is *place*. A geolibrary is distributed if its users, services, metadata, and information assets can be integrated among many distinct locations.

This report presents the findings of the Workshop on Distributed Geolibraries: Spatial Information Resources, convened by the Mapping Science Committee of the National Research Council in June 1998. The report is a vision for distributed geolibraries, not a blueprint. Developing a distributed geolibrary involves a series of technical challenges as well as institutional and social issues, which are addressed relative to the vision.

CHARACTERISTICS AND BENEFITS OF DISTRIBUTED GEOLIBRARIES

A wide variety of human activities could benefit from the services of distributed geolibraries. The activities include many for which the timely provision of information could minimize loss of life or result in more timely and effective use of existing information resources.

The contents of a distributed geolibrary are not limited to information normally associated with maps or images of the Earth's surface but include any information that can be associated with a geographic location. In this sense the vision thus extends far beyond the context of the National Spatial Data Infrastructure (NSDI).

New technological developments make it possible for people to gather data germane to their own needs more readily, extract data from online and other electronic repositories, develop the information products they need, use the products for decision making, and contribute their locally gathered geoinformation and derived products to libraries or other repositories. Developing the technical and institutional means to support incorporation of local knowledge into networked repositories presents a novel challenge.

Although many projects currently exhibit elements of the vision of distributed geolibraries, the lack of a clear statement of that vision impedes coordination and leads to duplication of effort. A clear statement can provide a sense of common purpose.

New technological initiatives such as the Next Generation Internet and Internet II are likely to provide extensions to Internet and World Wide Web (WWW) protocols and orders-of-magnitude increases in bandwidth. Many of these developments are expected to be relevant to distributed geolibraries.

THE NATIONAL SPATIAL DATA INFRASTRUCTURE

The vision of the NSDI as expressed by the Mapping Science Committee in 1993 (NRC, 1993) did not anticipate the

enormous impact and potential of the Internet and WWW. By emphasizing the problems of production of digital geoinformation, it underemphasized the importance of effective processes of dissemination to users. User communities are growing rapidly and are likely to grow even more rapidly if current difficulties associated with finding geoinformation on the Internet can be addressed.

Distributed geolibraries provide a useful framework for discussion of the issues of dissemination associated with the NSDI in addition to organization and access issues. The vision is readily extendible to a global context.

An essential component of a distributed geolibrary is a comprehensive gazetteer, linking named places and geographic locations. A national gazetteer would be a valuable addition to the framework data sets of the NSDI. These framework data sets are being coordinated by the Federal Geographic Data Committee (FGDC), which also has the responsibility for associated standards and protocols. Production and maintenance of the national gazetteer could be through the National Mapping Division of the U.S. Geological Survey (USGS) in collaboration with other agencies and could be an extension of the USGS's Geographic Names Information System.

CONTENTS, SERVICES, AND FUNCTIONS OF DISTRIBUTED GEOLIBRARIES

A distributed geolibrary would allow users (and computers) to specify a requirement, search across the resources of the Internet for suitable information, assess the fitness of that information for use, retrieve and integrate it with other information, and perform various forms of manipulation and analysis. A distributed geolibrary would thus integrate the browsing functions of the WWW with those of geographic information systems and related technologies.

In addition, a distributed geolibrary would support collaborative work, such as multidisciplinary research by teams, decision making by groups of stakeholders, and classroom projects by

groups of students. It would provide mechanisms for capturing the knowledge that results from such work and making it accessible to others as appropriate. It could also provide mechanisms for storing and archiving such knowledge.

Many important applications of distributed geolibraries are best located in the field, using portable systems and wireless communications. Delivery of services to the field is important in emergency management, agriculture, natural resource management, and many other applications.

The United States possesses vast archives of information that could be incorporated into distributed geolibraries and made accessible to users whose need for information is defined by geographic location. Linking much of this information to geographic location—in other words, to transform it to geoinformation—would be valuable within a geolibrary context.

Significant research problems will have to be solved to enable the vision of distributed geolibraries. Research needs include problems of indexing, visualization, scaling, automated search and abstracting, and data conflation. In addition, there are a variety of social and institutional issues that need further investigation. Research on these issues targeted to improve access to integrated geoinformation might be pursued by the National Science Foundation and other agencies sponsoring basic science, as well as by the National Mapping Division of the USGS, and the National Imagery and Mapping Agency.

ARCHITECTURE OF DISTRIBUTED GEOLIBRARIES

There are several alternative architectures for distributed geolibraries, including a single enterprise sponsored by a well-resourced agency, analogous to a national library; a network of enterprises with their own sponsors, analogous to a network or federation of libraries; and a loose network held together by shared protocols, analogous to the WWW.

INTELLECTUAL PROPERTY ISSUES

The development of distributed geolibraries will need to consider issues related to intellectual property rights. These need to be considered in the broader international debates about the nature of electronic information and databases as intellectual property. A distinction with respect to intellectual property rights needs to be drawn between raw data and knowledge works as they appear very differently from the perspective of the functions and services of a library. Strong arguments are presented for focusing distributed geolibraries on knowledge, rather than merely providing access to raw data.

ORGANIZATIONAL ISSUES

While traditional production of geospatial data has been relatively centralized, the vision of distributed geolibraries represents a broadly based restructuring of past institutional arrangements for the dissemination of geospatial data, one that is much more bottom-up, decentralized, and voluntary.

Many prototypes that include elements of a distributed geolibrary already exist, but it will take many years to realize the full vision, and it will be important to be able to measure and monitor progress. The vision of distributed geolibraries has distinct aspects that may not be addressed effectively by current programs aimed at digital libraries in general. The success of a distributed geolibrary is largely dependent on the ability to integrate information available about a place. That ability is severely impeded today by differences in formats and standards, access mechanisms, and organizational structures. Integration is a formidable problem for today's users of geospatial data.

1
Introduction

The Internet and World Wide Web (WWW) provide users with unprecedented access to information resources. In many ways they emulate the functions of traditional libraries, by making it possible to search and locate information using simple tools. But the potential is far greater in areas such as electronic commerce and in supporting new ways of finding information that go far beyond the services of the traditional library. One such possibility is the *distributed geolibrary*, the subject of this report. A distributed geolibrary would allow its users to search the resources of the WWW for information about a *place*,[1] to evaluate the information, and to retrieve and work with it as appropriate.

> A geolibrary is a digital library filled with geoinformation and for which the primary search mechanism is *place*. Geoinformation is information associated with a distinct area or footprint on the Earth's surface. A geolibrary is distributed if its users, services, metadata, and information assets can be integrated among many distinct locations. Chapter 2 develops a more detailed vision for geolibraries.

This report begins with a series of four examples to illustrate the range and importance of the practical problems that could be addressed by the services of distributed geolibraries. The following

[1] The term *place* is used throughout this report to refer to a location of interest on or near the Earth's surface. It might be a single point or an extended area or a volume above or below the surface; it might be defined by name or by coordinates, and it might be exact or ill-defined.

chapters discuss the full vision, social and institutional context, and steps that will need to be taken to make distributed geolibraries a reality. Because this is the first discussion of the topic, it falls short of a complete blueprint, and much more exploration will be needed. But this report is perhaps the first step in that direction.

Place is a common theme in many events, activities, emergencies, and issues. Terrorist acts like the World Trade Center bombing and natural disasters like Hurricane Andrew affect specific locations on the Earth's surface and call for relief efforts that must occur quickly and that are sharply focused in space. Accurate knowledge of the place at which an emergency occurs and of surrounding conditions is of critical importance in dispatching ambulances and other forms of relief. Place is important in learning about the world and in understanding its environment.

Distributed geolibraries are intended to provide new kinds of place-based information services that are not available from the traditional library or from the current WWW. The user of a distributed geolibrary should not be required to be an information retrieval expert, tò be proficient in computer technology, or to live in a metropolitan area. The distributed geolibrary envisioned in this report could be an information service for every American—for students and teachers, scientists, community members, government officials, business men and women, and families—by allowing ready access to available information about any place on the Earth's surface. The following hypothetical examples illustrate some of the potential uses and the critical importance of distributed geolibraries.

EXAMPLES

Emergency Response

A tanker truck carrying hazardous chemicals is traveling on the highway around a major metropolitan area. Just as the driver approaches a bridge his truck collides with the car in front of him. The truck flips, pinning both drivers inside their vehicles and

rupturing the tanker. From the debris a plume slowly rises from the chemical spill and is carried by the wind into the surrounding neighborhood. A liquid chemical drips over the bridge into the water below.

To deal with the emergency, metropolitan officials need to alert schools, residences, and businesses in the neighborhoods nearby. There has been a recent building boom, and new roads have been constructed. Local maps are out of date. Evacuations must be discussed and planned; routes need to be determined and reassessed; and the effects of weather on the plume need to be monitored continuously. Will it drift to the nearby airport as well? Meanwhile the spill must be contained, the traffic rerouted from the accident scene, and the way cleared for medical assistance. What human health hazards might be related to the contaminant? Hospitals and medical centers in the affected area must be put on alert.

Dealing with the potential contamination of the river requires considerable attention as well. What is the current rate of flow and level of the water? Who and what will be affected? Information is immediately required on towns, public and private sites, and beaches and harbors along the river. What access to these sites is possible? How can containment be achieved? The fast-running river passes many small communities and runs between two states. Data from many sources must be integrated and used in order for officials to deal with the effects of the accident. Other needs will emerge after the emergency is contained, such as dealing with the effects on wildlife habitat along the river and the fishing interests that flourish in the area.

But the immediate information needs are critical. Although emergency officials have access to their own local sources, they know some of their own maps are not current, so other data should also be checked. And the small towns along the river have limited information resources. The officials need services that allow them to access and browse available imagery, thematic maps, current public and private data resources, and even services available through commercial subscriptions. They need to reach other libraries

and online sites that specialize in key information, including contaminants. They need to identify personnel in other cities who have dealt with similar spills. In short, they need access to the best information available to cope with the emergency.

Information resources through distributed geolibraries could greatly assist rapid response to such emergencies and longer-term efforts aimed at prevention and mitigation. Moreover, it is important that information be available where it is needed most, which in many instances will be at the location of the emergency or in a local command center. The tools to access and work with information may have to operate in difficult environments using specialized field computers (palmtops, portables, or pen computers) and wireless communication. New sensors may be brought to the site, supplying data that will have to be integrated with existing data. Decision makers will want access to powerful aids for decision support and for rapid simulation of future scenarios.

Housing Relocation

A family is relocating to Southern California. They want to find a home in a suitable environment. They are concerned about earthquake hazards and want information that might help them avoid vulnerable areas and fault lines. After having identified several possible home sites, they further refine their search by excluding undesirable areas—such as high-crime districts or hazardous materials storage sites. They have read newspaper stories about brush fires. Has there been a history of such brush fires in any of the neighborhoods they are considering? They look at maps for the locations of churches, schools, shops, and parks. Special medical services are needed for one family member. What services are close? They consider distances to workplaces. They also worry about the wisdom of such a large investment. Will their home retain value? What are the neighborhood's economic trends?

The family wants to know about the *place* where they will live, work, and play. As responsible citizens they want to be informed about issues affecting their neighborhood. If such information is readily accessible, it could make a significant

difference in their choice of where to live. Today they might not have the resources, skills, or special education to find the answers to all of these questions, whereas most of this information would be available through the services of distributed geolibraries. In the future, however, they may be able to access information using wireless links directly to their vehicle as they explore potential neighborhoods.

Public Health

A researcher begins the task of analyzing the association of environment and disease in a particular urban area. She needs access to housing information and population characteristics, as well as health and medical histories in the geographic area of interest. She needs to examine health care facilities, types of buildings, disease rates, even summer heat fatalities, as well as environmental aspects, all over several decades. Incidents with contaminants and pollutants in the area must be located, assessed, and factored into her research. Finding the information will require searches through countless government institutions, media reports, and scientific journals.

She begins her work by visiting the local library; contacting responsible local, state, and federal agencies, talking with colleagues; and using search engines on the WWW. Finding the appropriate information, dealing with issues of confidentiality of health data, and putting the information into a form that can be integrated with other data about a given place can be time consuming; eventual success depends heavily on her background, technical training, and experience. Paradoxically, a request that can be expressed in very simple terms ("give me everything available about environment and disease *in this place*") turns out to be enormously and unreasonably complex, using the limited tools available today, and to consume the vast majority of the resources available to the project. Better tools for data access and management would allow more time to be spent on data analysis.

Natural Resource Planning

The year is 2010. More than 1,000 summer homes have been built within 10 miles of the boundaries of Yellowstone National Park, Grand Teton National Park, and the Bridger-Teton Wilderness Area. Numerous pets have been killed by grizzly bears, wolves, and coyotes, particularly in the early summer of 2009, when heavy snowpacks kept many wild animals from moving into the high country. The conflicts were capped by the deaths of a brother and sister, ages 7 and 8, following an attack by a grizzly bear, which was subsequently killed by wildlife authorities.

The National Park Service and the Fish and Wildlife Service are concerned about ever-increasing conflicts between wildlife and humans. Pressure from new residents and from ranchers has led to the death of 20 percent of the reintroduced wolves. Counties, once hungry for the economic growth brought by the construction of luxury summer homes, are now concerned about degradation of water quality and the demands of new residents that their assets be protected from wildlife. Fire management has become an increasing concern at multiple levels of government; officials recognize the need for frequent exposure of forests to fires in order to reduce fuel load, but with greatly increased private property near the forest they have found it increasingly difficult to allow fires to burn without risk to structures.

Local and federal agencies recognize the need to draw on common data resources that describe terrain, vegetation, and wildlife habitat in order to solve common problems of resource management. These data must be integrated across many different themes, topics, and disciplines and must be readily available to users needing to assess and plan effectively based on *place*.

The distributed geolibraries available to these stakeholders in 2010 allow them to assemble quickly information in the archives of the various levels of government, nongovernmental organizations, and citizen groups that are relevant to an issue centered at a particular *place* on the Earth's surface. Through distributed geolibraries, decision makers also may learn quickly what

information is not available elsewhere and therefore may need to be collected. Additional tools support the decisions and choices that need to be made. With these new tools, development of long-range plans that allow growth while minimizing conflicts with fire and wildlife is progressing after long delays. Several developments have now been completed in places where fire and wildlife conflicts are minimized and where drainage and sewage management have provided excellent protection of water quality.

A COMMON THEME

A common theme in these examples is the current inability to locate and integrate information quickly and simply based on *place*. Although place is the definitive element in many issues, it is currently easier to find information about a named individual, an agency, or a field of scientific knowledge than about a place on the Earth's surface. This report explores opportunities that will improve our ability to find, access, integrate, and use information by exploiting the technologies of the Internet, the WWW, geographic information systems, and digital computers.

Finding 1
A wide variety of human activities could benefit from the services of distributed geolibraries. They include many where the timely provision of information could minimize loss of life or result in more timely and effective use of existing information resources and others where the costs of bad decisions could be avoided.

Distributed geolibraries could provide information services directed specifically at the needs of communities. In a speech given at the Brookings Institution on September 2, 1998, Vice President Gore argued that increased public access to information through mechanisms such as those discussed in this report will put "more control, more information, more decision-making power into the hands of families, communities, and regions, to give them all the freedom and flexibility they need to reclaim their unique

place in the world." The services of distributed geolibraries that are discussed and elaborated in this report could enhance education, improve the quality of day-to-day living, and provide economic benefits. They could support scientific research by furnishing new tools for search, analysis, data fusion, and visualization. They could provide the means by which officials cope with emergencies, address issues of health and social services, troubleshoot crime, and accomplish urban planning. They could help provide economic benefits by enabling people to research, manage, market, and grow their business ventures.

Many of the components of distributed geolibraries already exist or are being developed, and many existing WWW sites offer some limited form of distributed geolibrary services. This report goes beyond the present to articulate a vision of what might be, with the objective of providing a common target and of pulling disparate threads together into a unified effort to achieve that vision in the not too distant future.

2

A Vision for Distributed Geolibraries

RECENT DEVELOPMENTS

The past two decades have seen rapid developments in information technology. Hardware components have become smaller and more powerful, enabling the development of the personal computer and bringing the ability to process information to field environments that are far removed from the office and desktop. Software has grown more sophisticated, empowering individuals with little technical training to make effective use of computers in ways that would have been inconceivable 25 years ago. Developments in wireless communications allow networked access virtually anywhere. Most recently, applications of the Internet and World Wide Web (WWW) have captured the popular imagination and spawned entire industries of electronic commerce and information dissemination.

These developments have in turn driven massive changes in the way society disseminates and accesses information of various types. The role that information plays in everyday activities is changing, as people come to rely on access to up-to-the-minute information on weather, markets, politics, and entertainment via the Internet. Changes seem especially challenging and profound in the area of information that is tied or related to a geographic *place*, that is, a location at or near the surface of the Earth. Millions of people access such WWW sites as MapQuest (*www.mapquest.com)* or Microsoft's Terraserver

(*www.terraserver.com)* each day, which offer maps, driving directions, satellite images, and other forms of raw or processed information and related services (see Appendix D for examples). Similar changes are reflected in the proliferation of geospatial data clearinghouses, digital spatial data libraries, geographic information system software, and new high-resolution imaging satellites.

Several factors help explain the high level of interest in the Internet and WWW as technologies for disseminating these particular types of information and related services. First, the methods of storage and dissemination of traditional products—paper maps, atlases, and photographic images—are cumbersome in comparison to digital data products and often require special cabinets and awkwardly shaped shipping packages. Digital methods make it as easy to store or send a map as it is to handle text. Second, geoinformation is often related to a specialized interest, and it may be hard to justify maintaining an extensive collection in a local library or bookstore; the WWW is ideally suited to the distribution of such information in response to specialized needs because the costs of maintaining a server are low, and universal access to the Internet means that only one server is needed. Finally, geoinformation needs to be timely, but it can take years for a paper map to be produced, printed, and disseminated; the WWW allows users to access information as soon as it is posted.

At the same time there are potential disadvantages to use of the WWW as a mechanism for storing and disseminating geoinformation that will have to be addressed. Little of the information now available via the WWW has been subjected to the mechanisms that ensure quality in traditional publication and library acquisition: peer review, editing, and proofreading. There are no WWW equivalents of the library's collection specialists who monitor library content. But it is easy to be misled into believing that quality control problems of the WWW and distributed geolibraries are somehow different from conventional ones. Users of distributed geolibraries will tend to trust data that come from

reputable institutions, with documented assurances of quality, and to mistrust data of uncertain origins, just as they do today.

A common theme in all of these efforts to exploit the Internet and the WWW has been the enabling role of technology; many people with an interest in geoinformation and an awareness of the potential of the WWW and related technologies like the Java programming language have begun exploring their use. Five years after the first explosion of interest in the WWW is an appropriate time to pause and ask some basic questions:

- Is there a vision that drives the efforts to build clearinghouses and other WWW-based access and dissemination mechanisms for geoinformation?
- What need are these efforts satisfying, from the perspectives of the users and producers of geoinformation and the providers of related services?
- What problems impede progress, and on what problems should efforts be expended?
- What high-priority research needs exist?
- How should public resources best be expended, and what new forms of collaboration are needed?

The Mapping Science Committee convened a workshop[1] in June 1998to explore these issues. The Workshop on Distributed Geolibraries: Spatial Information Resources was designed to explore long-term visions of how ongoing activities may evolve, to explore possible development strategies, and to identify common needs (see Finding 2). Workshop participants were selected to represent a number of communities with interests in these issues:

[1]The workshop (and this report) focused on the discovery, access, integration, and use of geoinformation. Other technical issues (e.g., archiving, quality control and assurance, standards development, telecommunications and computational capabilities), although critical in the development of the distributed geolibrary concept, were not extensively considered.

experts in dissemination of geoinformation; leaders of current activities; specialists in the relevant technologies; and specialists in the associated institutional, legal, social, and economic issues. A list of participants is provided in Appendix A.

Finding 2
Although many projects currently exhibit elements of the vision of distributed geolibraries, the lack of a clear statement of that vision impedes coordination and leads to duplication of effort. A clear statement can provide a sense of common purpose.

Prior to the workshop, participants were asked to contribute a "white paper" on issues they found relevant to the topic. These papers, which provided useful background to the meeting, are listed in Appendix B and are available on the WWW.

This report was prepared by the panel that organized the workshop (a list of panel members appears in the beginning of this report). Thus, it reflects the consensus of the panel, regarding the discussions that took place at the workshop, the issues that arose there and in the white papers, and the workshop's broader context.

The workshop did not attempt to bound the scope of distributed geolibraries precisely, and even if that were possible it would have been unreasonable to expect it in a workshop of such limited duration. Many basic questions remain unanswered, and this report should be read as a first effort in this area and as a stimulus for further work and discussion, rather than as a precise blueprint.

The workshop participants were almost entirely from the United States, and this report necessarily adopts a U.S. perspective. Nevertheless it is hoped that it will be read by non-U.S. researchers and developers interested in distributed geolibraries and that it will help to achieve a greater degree of convergence in research and development at the international level.

A LIBRARY VISION

The organizers of the workshop chose to frame the discussion by reference to the functions, services, and institutional arrangements of the library, for two major reasons: first, to engage the library community, with its long experience in providing access to information, in the development of a vision for a new kind of library and, second, to provide a familiar and concrete starting point for the discussion. It is possible that libraries will be the principal means whereby citizens gain access to the services of the distributed geolibraries of the future; it is also possible that libraries will play no significant part in that process.

The metaphor of the library is powerful because it immediately suggests a number of important issues. For example, one way to think of a library is as a storehouse of the intellectual works of society, and millions of people from all walks of life have contributed works to our current library system. Can we expect to see a similar diversity of contributors in the distributed geolibraries of our future? What incentives are needed to motivate people to make their works accessible? If a library exists to serve a community, its first responsibility should be to provide the information needed by the community. How important is geospatial information about the community itself, produced perhaps within the community, compared to information about areas outside the community perhaps produced by others? Will a local geolibrary, responsible to a local community, acquire and make available very different works and databases than a university-based geolibrary, state geolibrary, federal agency geolibrary, or a private geolibrary?

There are many types of libraries and much variation in the functions they perform. Some of the comments in this report refer to all types of libraries, and some are more appropriate for the research library, the institution maintained by a university, or similar organization for the use of its community of scholars. In general, it is the research library that provides the model of services discussed in this report.

However, the metaphor of the library should not be taken too far, and not all aspects of the operation of a library will be useful in envisioning distributed geolibraries. Many of these will be generic and of no specific relevance to the geoinformation that is the focus of distributed geolibraries. Such issues have already been discussed at length in the library and digital library literatures, and no attempt is made to replicate those discussions here. For example, it is assumed that distributed geolibraries will need to address issues of archiving and preservation (particularly serious issues given the rate of technological change in the digital world), but these are generic to all libraries and are not discussed at length in this report.

DEFINING A DISTRIBUTED GEOLIBRARY

Three ideas help to define the concept of a distributed geolibrary: it is *distributed*, modeled on the concept of a *library*, and concerned with information about the *Earth*. The next three sections discuss these ideas in detail and build an outline of a vision for distributed geolibraries.

A Distributed Library

The term *distributed* refers to the locations of the physical and functional parts of the library and the locations of its users. In a traditional library the various stages of putting useful information into the hands of users occur largely in one place, in the physical structure known as the library. Books arrive in an acquisitions department; they are cataloged by specialists employed by the library in a cataloging department, placed on shelves within the library in locations designed to make it easy for patrons to browse through holdings on similar topics, retrieved by librarians and users, and signed out of the library at the circulation desk operated by a circulation department. Because these functions occur in one institution, it is sometimes difficult for an observer to separate

them and difficult to distinguish the functions of the library from its physical assets.

In today's digital world it is possible for functions to occur in multiple locations, held together and coordinated by communications networks like the Internet. Catalog staff may work in locations far removed from the reference librarians who eventually use the catalog to help users find the information they need. Moreover, today's technology is advancing to the point where patrons (or users) can employ library services to combine data sets located in different places. For many purposes the Internet provides almost infinite connectivity, such that a user may conceive of a single database that is in reality distributed over many different servers under different jurisdictions. Users have the option of processing data on their own computers or sending data to remote locations where processing capabilities are more powerful. Wireless technologies provide for communication to virtually everywhere, and computing technology can now be packaged into electronic units that are readily transportable and in some cases wearable.

Libraries have responded to this new networked environment by establishing coordinated, collaborative, and multi-institutional relationships. The library building no longer houses all of the services it provides to its users; instead, the institution of the library obtains those services in whatever ways maximize effectiveness and minimize costs, by using resources in the building or from a myriad of sites distributed around the globe.

Traditionally, libraries have made a clear distinction between *general* and *special* collections, using the latter term to refer to assets that need special treatment or that are unique in some way to a particular library, such as the papers of a particular literary or scientific figure. Maps and images form special collections in many libraries, in part because they are difficult to handle and in part because much of the collection may be unique. The transition to a digital world will mean that many of the difficulties of *handling* special media disappear, allowing such collections to become part of a library's information mainstream

(although working with maps and images will always demand specially designed interfaces and large monitors because of their visual content and broad bandwidth and powerful processors to deal with voluminous data). But the *uniqueness* of the special collection will become increasingly important in the digital world, in which any item in any collection is potentially accessible from anywhere.

In this report the term *custodian* refers to the person or agency responsible for maintenance of a given data set. The custodian may be far removed from the server on which the data set is mounted and from which it is disseminated, but nevertheless it is the custodian who holds the definitive version of the data and updates it to account for changes. The custodian may have some form of responsibility for quality—for example, the custodian may decide which data are to be acquired and held based in part on quality or may provide assurances of quality to users. The function of a custodian is different from that of a *repository* or *archive*, which is where data are preserved in static form.

Geoinformation

Geoinformation is information that is specific to some part of the Earth's surface or near surface. It includes maps, of course, which abstract and present information about the locations of phenomena on the surface; it also includes images from the air or space (aerial photos or remotely sensed images) that capture the appearance of the surface using energy (either visible or invisible) radiated from it in some part of the electromagnetic spectrum. Such data were earlier defined as *geospatial*. In addition, geoinformation includes the contents of guidebooks, reports on specific areas, data sets with a geographic dimension, and any other information assets that serve to differentiate one geographic area from another. Finally, it includes information about the atmosphere above the surface, the geology below the surface, and the oceans that cover two-thirds of the planet's surface.

All of these information assets are characterized by having some form of associated geographic *footprint*, a boundary defining

the geographic extent of the information, which is the defining characteristic of geoinformation as the term is used here. A map sheet has a footprint defined by its edges, whereas a guidebook to Moscow has a footprint of the city limits (or the city and the surrounding region). A photograph might have a footprint, defined as the area shown in the photograph; a piece of music (George Gershwin's "An American in Paris," for example) might also be associated with some particular location on the Earth's surface. Moreover, *the footprint provides a useful way of finding information*. Just as author, subject, and title are ways of finding information assets in a traditional library, so the footprint of geoinformation gives the library the ability to identify all those assets that fit a given geographic query. For example, if information assets in the library had a footprint, it would be possible to identify those assets relevant to a user wanting information on the state of Missouri, or the Caspian Sea, by determining whether the footprint of the asset matched the footprint of the query in whole or in part. It would be possible to ask the library to provide all available information about a given place that is relevant to a defined need, in other words "everything relevant about *there*."

While the space of a search based on author or subject is discrete, geographic space is continuous and multidimensional, and there is no limit to the number of distinct, unique footprints that exist. Any degree of overlap is possible between a footprint and a query, making search by place inherently more complex than search by other keys. Geographic location is sometimes recorded in the subject fields of library catalogs (for example, the Melvyl catalog of the University of California library system includes a place-related subject in about 30 percent of all records), and it is included in the Dublin Core standard (*purl.org/dc*). But distributed geolibraries would prioritize place as the primary key and thus would require that footprints be explicit in all cases.

Two distinct methods are available for specification of footprints. An area of interest may correspond to one or more *place names*, or recognized terms for describing location. Alternatively, the area may be defined by one or more bounding

coordinates, in some recognized system such as latitude and longitude. To be compatible, the two methods require the services of a *gazetteer*, or an index that relates named places to coordinates. Gazetteers are commonly used to index atlases, though as the name suggests they typically include only places whose names have some level of official recognition.

The issues surrounding place as a search key are to some extent similar to those surrounding time, or date. All of the examples in Chapter 1 require search by place, in many cases qualified by relevant intervals or points in time; perhaps it is possible to devise parallel examples that would require search by time, possibly qualified by place, to motivate the development of *chrono*libraries. Similarly, an important but less compelling case can be made for a three-dimensional approach to space, based on examples of data that relate to points substantially above or below the Earth's surface.

Spatial keys are not unique to geoinformation, and there are parallels to other domains that may be useful and informative in the development of distributed geolibraries. For example, the Hytime hypermedia document structuring language (Newcombe *et al.*, 1991) includes standards for specification of spatial windows in arbitrary coordinate systems within documents.

Geoinformation can be cumbersome for the traditional library because it comes in many forms, on different media, and because there is no simple basis for cataloging it. Instead, map libraries and other stores of geoinformation have had to maintain expensive and highly trained staffs to help users navigate through their information resources, and users have had to look to numerous sources to meet their geoinformation needs. Users of geoinformation were often highly trained experts, knowledgeable about sources, data quality, acronyms, and other tools of the geoinformation trade. In short, there has been no way for an average person to address a library with the query "tell me everything you have about that *place* that is relevant to me." Yet such queries are common and immensely important to a wide

range of human activities, as the examples in the opening chapter illustrate.

Although it is helpful to think of a distributed geolibrary as a container of the digital equivalent of maps, that metaphor may also be unduly limiting. Geoinformation is not restricted to information that is static, or two-dimensional, but includes information on the dynamic processes and changes happening at a place, and three-dimensional data about the atmosphere and subsurface. But as noted earlier, the two horizontal dimensions are most likely to be the basis for search, possibly refined by time and the vertical dimension.

Characteristics of a Distributed Geolibrary

One way to think about a geolibrary (in a world of paper documents) is to imagine walking into a library building and being confronted not with a card catalog, or its modern digital equivalent, but with a giant physical globe. Suppose what is needed is information about a particular part of Patagonia, the southern extremity of Argentina, for a project on Charles Darwin, who visited Patagonia, or on the people of Welsh descent who live there, or on the works of author Bruce Chatwin, who wrote about his travels there. The library user finds Patagonia on the globe, points to it, and asks a nearby librarian about the relevant assets of the library. Some minutes later the librarian produces a list of those assets, with enough information to allow the user to evaluate their importance to the project. After the user narrows the list, the librarian disappears again, to return with the requested holdings.

Several aspects of this concept ensure that it has remained in the realms of fiction for as long as libraries have existed. Some aspects are technological. There is no way to build a physical globe that can be repositioned at will or magnified on demand to display greater and greater detail. Zooming would need to be possible over several orders of magnitude; a large globe might reasonably be expected to show features on the Earth's surface that are 10 km in size, including large lakes and large cities, but not features as small as a neighborhood; but the user of a geolibrary

might well want to consider a single city block, which requires a resolution finer than 10 m, or a factor of 1,000 finer than the initial coarse view. Such resolutions are increasingly common in geospatial data.

In addition to resolution, a physical geolibrary would be difficult to build because many of its users would not be able to find their areas of interest on the globe. Not every user would be able to reposition and zoom to identify his or her own neighborhood, without the assistance of an expert. There are not enough resources to support the necessary expert librarians and no way to transform automatically a specified location into a list of assets. Finally, there is no way to shelve the many different types of information so that they can be easily retrieved and so that two sources of information on similar topics or areas are located near each other in the library. In other words, a physical geolibrary cannot be built.

In a digital world, however, all of these objections disappear, apparently without exception. It is possible to present the digital library user with a picture of a globe; search for locations by name, address, or any other suitable and convenient method; allow repositioning and zooming; search distributed archives for information assets whose footprints match the query, present them to the user in sufficient detail to permit evaluation; and deliver them for further examination and analysis. But although a geolibrary is possible in principle, there are countless technical, practical, economic, and institutional problems that will have to be overcome. Moreover, it is unclear how a geolibrary would deal with issues of intellectual property and how it could be paid for and whether the costs would be outweighed by the benefits. These issues are explored in greater detail in Chapter 3.

A distributed geolibrary would provide a much more sophisticated and powerful layer of services above the Internet and the WWW (Figure 2.1). The Internet provides the means of communication between computers, using the TCP/IP standard. The WWW is supported by the Internet, providing services that allow any user to access information provided by any server. But

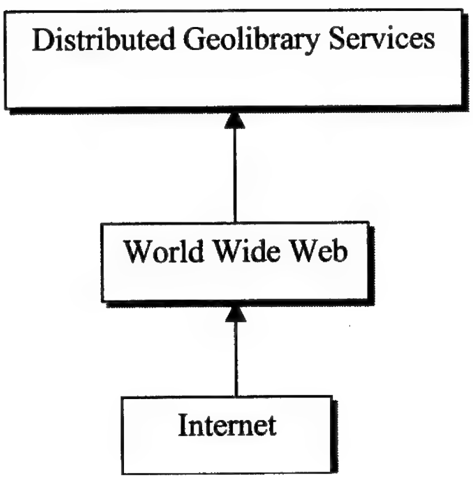


FIGURE 2.1. Distributed geolibraries as a third layer of services above the WWW and the Internet.

the combination of the two technologies falls far short of the services of a distributed geolibrary:

- The WWW does not have an equivalent of the library's carefully constructed catalog of assets. Search services such as AltaVista, Yahoo, and eBLAST that substitute for the services of a WWW catalog are crude imitations of the sophisticated skills of information abstraction possessed by the professional librarian.
- The number of WWW servers is now on the order of 10^7 and increasing rapidly. Even the most powerful of today's search engines can access no more than one-third of what is currently available, and this proportion decreases daily (National Public Radio report dated 3 April 1998 in a recent article in *Science*).
- Footprints and other essential information are not normally present in WWW information resources, and there are limited tools to look for them or to determine them automatically. Chapter 4 of this report discusses existing efforts to develop some of these services, and Appendix D includes examples of current projects

and sites that offer some of the services of distributed libraries, such as the University of California's Alexandria Digital Library.

- Users must rely on personal knowledge to find sites that contain needed information assets and must learn the specific protocols used by each site.
- There are no generally available services for combining information from multiple sources or for support of analysis, visualization, and interpretation of geoinformation by the user, although such services have been developed in limited contexts, including U.S. Department of Defense applications.

In other words, a distributed geolibrary would constitute a level of services above those provided by the Internet and the WWW, geared to specific user needs. Distributed geolibrary services offer the potential for more intelligent organization and access, for the creation of new knowledge through analysis of raw data, and for the solution of practical problems. As such, distributed geolibraries are one of a number of new types of Internet services that exploit previously impractical ways of organizing and presenting information.

DISTRIBUTED GEOLIBRARIES AND THE NATIONAL SPATIAL DATA INFRASTRUCTURE

"The National Spatial Data Infrastructure is the means to assemble geographic information[2] that describes the arrangement and attributes of features and phenomena on the Earth. The infrastructure includes the materials, technology, and people necessary to acquire, process, *store*, and *distribute* such information to meet a wide variety of needs" (National Research Council, 1993, p. 2, emphasis added). The concept emerged in the

[2]The term *geographic information* here is synonymous with *geospatial data* as defined in Chapter 1. But as noted earlier in this chapter, many additional types of information qualify as geoinformation by virtue of having a geographic footprint.

early 1990s in response to a number of potentially critical trends that were affecting the nation's supply of geospatial information and related services and institutions:

- Budgets in the federal public sector were declining and were no longer able to meet the nation's growing needs for high-quality, current geospatial data at minimal cost to users.
- Improved and cheaper mapping technology was empowering local and state governments to produce their own geospatial data to meet local needs and stimulating a growing private-sector industry.
- Advances in digital technology were making it possible to integrate and analyze geospatial data and support decisions in more powerful ways.

The Mapping Science Committee's report *Toward a Coordinated Spatial Data Infrastructure for the Nation* (National Research Council, 1993) and the efforts of many other individuals and agencies led in 1994 to Executive Order 12906, by which President Clinton ordered the development of the National Spatial Data Infrastructure (NSDI). Since then, several other committee reports and extensive efforts by the Federal Geographic Data Committee (FGDC), National States Geographic Information Council (NSGIC), National Association of Counties (NACO), and other groups have refined the concept of the NSDI and demonstrated its power and effectiveness (Tosta and Domaratz, 1997; Moeller, 1998; Rhind, 1999).

Finding 3
The contents of a distributed geolibrary are not limited to information normally associated with maps or images of the Earth's surface but include any information that can be associated with a geographic location. In this sense the vision extends far beyond the context of the NSDI.

When the NSDI was defined in 1993, few users or producers of geospatial data made much use of the Internet, and

the WWW was virtually unknown; the first popular browser, Mosaic, was released by the National Center for Supercomputer Applications early that year. Although there was much emphasis on digital geospatial data, the primary method of dissemination was by magnetic tape; there were virtually no digital online catalogs of geospatial data and no methods for searching for data across computer networks. Moreover, since most useful geospatial data were produced by a small number of federal agencies, there was little problem locating the appropriate source. WAIS (Wide Area Information Service) was the first of several network-based technologies that rapidly changed the nature of geospatial data dissemination over the next few years. Today, applications on the WWW have grown into an enormously successful tool, and have had a profound impact on the entire environment for geoinformation acquisition (National Academy of Public Administration, 1998). At the same time, the WWW has presented a growing problem in its inability to deal effectively with the problems of discovering what geoinformation exists and locating an appropriate source, as the number of potential suppliers has mushroomed.

Finding 4
The vision of the NSDI as expressed by the Mapping Science Committee in 1993 (National Research Council, 1993) did not anticipate the enormous impact and potential of the Internet and WWW. By emphasizing the problems of production of digital geoinformation, it underemphasized the importance of effective processes of dissemination to users of geoinformation. User communities are growing rapidly and are likely to grow even more rapidly if the current difficulties associated with finding geoinformation on the Internet can be addressed.

This report and related efforts in general can be understood therefore as an updating of the Mapping Science Committee's concept of the NSDI in the era of the WWW. In organizing this effort and producing this report, the committee is expressing its

view that the WWW has added a new and radically different dimension to its earlier conception of the NSDI, one that is much more user oriented, much more effective in maximizing the value of the nation's geoinformation assets, and much more cost effective as a data dissemination mechanism. Distributed geolibraries reflect the same basic thinking about the future of geospatial data, with its emphases on sharing, universal access, and productivity but in the context of a technology that was not widely accessible prior to 1993.

Finding 5
Distributed geolibraries provide a useful framework for discussion of the issues of dissemination associated with the NSDI. The vision is readily extendible to a global context.

The NSDI fits well with the description of infrastructure provided by Star and Ruhleder (1996, pp. 111–112):

> "It is both engine and barrier for change; both customizable and rigid; both inside and outside organizational practices. It is product and process.... With the rise of decentralized technologies used across wide geographical distance, both the need for common standards and the need for situated, tailorable and flexible technologies grow stronger."

Their defining dimensions of infrastructure provide useful guidance to the development of distributed geolibraries: they would be *embedded* in other structures, social arrangements, and technologies; their *reach or scope* would extend beyond a single site or practice; their procedures would be *learned as part of membership* of an organization or group; they would be *linked with conventions or practice* of day-to-day work; they would be the *embodiment of standards* and would *build upon an installed base*; and they would be *visible on breakdown*, since we would be most aware of them when they failed to work.

DISTRIBUTED GEOLIBRARIES AND DIGITAL EARTH

Distributed geolibraries bear a strong resemblance to certain aspects of the concept of Digital Earth, a concept that was defined by Vice President Gore in January 1998 and summarized in a speech given in Los Angeles. The vision is aptly summarized in the following extract:

> "Imagine, for example, a young child going to a Digital Earth exhibit at a local museum. After donning a head-mounted display, she sees Earth as it appears from space. Using a data glove, she zooms in, using higher and higher levels of resolution, to see continents, then regions, countries, cities, and finally individual houses, trees, and other natural and man-made objects. Having found an area of the planet she is interested in exploring, she takes the equivalent of a 'magic carpet ride' through a 3-D visualization of the terrain. Of course, terrain is only one of the numerous kinds of data with which she can interact. Using the system's voice recognition capabilities, she is able to request information on land cover, distribution of plant and animal species, real-time weather, roads, political boundaries, and population. She can also visualize the environmental information that she and other students all over the world have collected as part of the GLOBE project. This information can be seamlessly fused with the digital map or terrain data. She can get more information on many of the objects she sees by using her data glove to click on a hyperlink. To prepare for her family's vacation to Yellowstone National Park, for example, she plans the perfect hike to the geysers, bison, and bighorn sheep that she has just read about. In fact, she can follow the trail visually from start to finish before she ever leaves the museum in her hometown.

> She is not limited to moving through space, but can also travel through time. After taking a virtual field-trip to Paris to visit the Louvre, she moves backward in time to learn about French history, perusing digitized maps overlaid on the surface of the Digital Earth, newsreel footage, oral history, newspapers and other primary sources. She sends some of this information to her personal e-mail address to study later. The time-line, which stretches off in the distance, can be set for days, years, centuries, or even geological epochs, for those occasions when she wants to learn more about dinosaurs."

Digital Earth is also the title of a project[3] of several years' standing at NASA's Goddard Space Flight Center, which also contains elements of the Vice President's vision. It is also associated with a plan to place a satellite (tentatively named "Triana") between the Earth and the Sun to deliver real-time images of the sunlit Earth to a global audience.

Like distributed geolibraries, Digital Earth is about making use of the vast but uncoordinated masses of geoinformation now becoming available via the Internet and about presenting it in a form that is readily accessible to the general user. Like distributed geolibraries, its central metaphor for the organization of information is the surface of the Earth and *place* as a key to information access. In a similar vein the U.S. Geological Survey is exploring the Earth's surface as the organizing metaphor for public access to its data resources, and similar ideas are surfacing in other agencies (see Appendix D).

Learning about places on the Earth is a strong theme in Vice President Gore's vision for Digital Earth and a strong motivation for distributed geolibraries. While the prevailing metaphor for human-computer interaction is the office or desktop, that metaphor may not be particularly helpful in organizing information about the Earth. Instead, access to a distributed geolibrary could be through the visual metaphor of the Earth's surface itself; a student

[3] *http://holodeck.gsfc.nasa.gov/digitalearth/digitalearth.html*

interested in Thailand would manipulate a globe on screen until it centers on Thailand and then zoom in for more detail, as in the Digital Earth vision. Distributed geolibraries might make a useful contribution to the educational opportunities of digital libraries, as outlined, for example, in previous reports on digital libraries for science, mathematics, engineering, and technical education (see Corportaion for National Research Initiatives, 1998; National Research Council, 1998).

The library service model that underlies the concept of distributed geolibraries provides a useful way of structuring discussion and of thinking about the resources and research that will be needed to make the vision a reality. Chapter 3 discusses some of the societal and institutional challenges to realizing distributed geolibraries. Addressing many of these policy issues is crucial to creating a conducive atmosphere for considering the potential services and functions of distributed geolibraries (see Chapter 4) and the technical developments needed to build distributed geolibraries (see Chapter 5).

3
The Distributed Geolibrary in Societal and Institutional Context

Implementation of a distributed geolibrary presents a host of challenges, ranging from the technical to the societal and institutional. The latter are discussed in this chapter; technical issues are discussed in Chapter 4.

The policy challenges presented by distributed geolibraries include the following:

- What are the legal, ethical, and political issues involved in creating distributed geolibraries? What problems must be addressed in the area of intellectual property rights? How will these issues affect the technical development of distributed geolibraries?
- Who will pay for the creation and maintenance of distributed geolibraries? What components might be in the public domain versus those provided by the commercial sector?

This chapter addresses many of these issues from the perspective of geoinformation at the local level, how distributed geolibraries might build off the library model (and how traditional libraries have addressed or handled some of these societal and institutional issues), and some of the additional issues introduced by the digital context of distributed geolibraries. These issues are not necessarily unique to distributed geolibraries as many have been discussed extensively within the context of recent digital library programs. The intention here is not to review or paraphrase excellent surveys of the social context of digital libraries, such as that of Borgman *et al.* (1996), which readers interested in a broader perspective should consult.

LOCAL FOCUS

Five years ago discussions regarding geospatial data in the United States focused on the rapidly increasing use of such data throughout society and the need to create a more formal infrastructure to coordinate geospatial data coverage across the nation, minimize redundant data collection at all levels, and create new opportunities for use throughout the nation (National Research Council, 1993). Much has been accomplished. Concepts such as metadata standards, standard framework databases, and thematic databases have been developed and pursued (see *www.fgdc.gov*). The federal government in cooperation with state and local governments has been and continues to be well positioned to lead the development of the basic concepts and public domain databases upon which the NSDI is being built.

The NSDI now involves many stakeholders as a result of activities over the past five years. Its basic data will be assembled from diverse institutions throughout the nation, with institutions contributing those parts that are most relevant to their roles (Tosta and Domaratz, 1997; Moeller, 1998; Rhind, 1999). At the core of this vision is the concept of local generation of geoinformation. Geoinformation is inherently local in nature and of greatest importance to those in that local area. It makes sense that the tens of thousands of units of local governments in the United States understand their own geoinformation assets and needs far better than do higher levels of government.

New developments in technology make it possible for local people to gather local data germane to their own needs more readily, extract data from online and other electronic repositories, develop the information products they need, use the products for decision making, and contribute their locally gathered geoinformation and derived products to libraries or other repositories. Developing the technical and institutional means to support incorporation of local knowledge into networked repositories presents a novel challenge.

Stakeholders across the nation are beginning to think and act around more common visions for the NSDI. A library service model

provides an initial way to consider the organizational and institutional arrangements for finding and accessing the geoinformation assets and digital products being generated by numerous stakeholders across the nation.

LIBRARY CONSIDERATIONS

The Library as an Institution

In considering possible institutional arrangements for distributed geolibraries, we begin with the assumption that libraries are social institutions that will continue to change but will not be made obsolete by the advent of electronic publishing. Indeed, distributed geolibraries and digital libraries in general will complement the traditional activities of libraries and related institutions. Libraries respond to many complex societal needs. They are used for research, teaching, self-learning, and entertainment. They serve as social and activity centers for many communities, whether these be small towns, neighborhoods, or institutions. The opportunities that libraries provide range from learning about practical matters to exploring science, art, history, or literature for the sheer pleasure of doing so. They are places for children to learn how to read and places for disadvantaged members of communities to seek solutions and solace (Crawford and Gorman, 1995, p. 118). The library system serves as a repository and by doing so preserves most aspects of our culture. Libraries range from small to large, urban to rural, and public to private but cooperate through a common professional culture and set of procedures, sharing information for mutual benefit. In short: "libraries exist to acquire, give access to, and safeguard carriers of knowledge and information in all forms and to provide instruction and assistance in the use of the collections to which their users have access" (Crawford and Gorman, 1995, p. 3).

Libraries have incorporated information technologies in all aspects of library services. Most recently, libraries have embraced network-based programs that support collaboration among

institutions and the sharing of resources. In addition, consortia have been established on state, regional, and library-type bases throughout the United States to share information, negotiate licenses, engage in collection development, and for many other purposes. A useful distributed geolibrary of the future will need to participate in these activities as an entity that will accumulate, make available, and conserve electronic carriers of georeferenced knowledge.

Economic Considerations

Existing public libraries do not buy most books or subscribe to most magazines or journals, yet they are highly valued by the estimated two-thirds of American adults who use them (Crawford and Gorman, 1995, p. 127). A typical robust public library will lend out 10 items per person per year based on the population served by the library and will answer two questions per person per year for its service population. Typical circulation of a robust library is twice its content (i.e., a library with a collection of 1 million volumes will lend out 2 million volumes during the year). In-library use of volumes in poor and rural communities often exceeds circulation, and in-library use at academic libraries often exceeds circulation by two to three times.

Public libraries provide these high use and service rates at a cost of approximately five cents per day per capita for their service population, while public libraries in economically healthy areas aspire to 10 cents per day per capita as a reasonable starting point for funding a robust library (Crawford and Gorman, 1995, p. 139). These expenditures appear to be a bargain for the access and services provided, and any proposal for supplanting current library services with electronic services would need to compare costs realistically.

Conversely, would an electronic digital library be available to at least the two-thirds of American adults who currently use existing libraries? Would it serve children and the disadvantaged to the same extent or greater than existing library facilities and resources?

There is an economic conundrum that in the face of a proportionately higher demand some communities might not have the available resources to support distributed electronic delivery services, even though the delivery technology is dropping in price. In terms of distributed geolibraries, this may be an issue, as a recent survey of public libraries in Colorado (Gayon, 1998) indicates that rural libraries receive a larger than expected proportion of requests for geographic information (maps, images, and digital data).

Libraries have the effect, although not a priority purpose, of introducing library users to works, authors, and publishers. Libraries thereby serve the economic function of creating markets for intellectual works. Would a geolibrary have the same effect? These are some of the institutional questions that will need to be addressed as the technological capabilities for distributed geo-libraries are built over time.

Distributed Geolibraries and the Existing Library Institution

Might distributed geolibraries develop as part of existing library arrangements or complement them? Although the possibility exists that distributed geolibraries might develop in tandem with libraries and be interconnected with them, the duplication of all the roles of libraries in a new institutional environment would make little sense. A useful analysis of these issues is presented by Hawkins (1994) in the context of digital libraries. Indeed, the way distributed geolibraries evolve will depend in large part on access to resources in existing library institutions.

Some of those things that traditional libraries have never been able to do well might be better done by digital means. One of these functions might be the provision of access to geoinformation. The size and shape of the sheets on which paper maps are produced often depend on the information or the story that the cartographer is attempting to convey graphically, the scale required to present information adequately, and the shape of the geographic area being addressed. Owing to the wide variability in map sizes and the nonstandard placement of information on them, the classification, cataloging, and storage of maps have been far

more problematic for librarians than handling books, journals, magazines, and recordings. Thus, in some instances, maps may be ineffective uses of print on paper, and many maps might be better represented, accessed, and used in digital form.

Thus, the advent of distributed geolibraries is likely to alter the relative advantages of paper and electronic map production. Paper map collections in libraries are unlikely to be completely eliminated. Because of the increasing user friendliness of mapping software and the ready availability of digital geoinformation, the ability to produce sophisticated maps and communicate through them is now available to many more people. As a result, we may witness substantial increases in both paper and digital maps that may be of interest to members of communities and made available in their local libraries.

Although a geolibrary is defined earlier in this report as digital in nature, any practical or useful geolibrary from an institutional perspective will need to be able to accommodate a multiplicity of forms for conveying knowledge. The various means and forms for conveying geographic knowledge each have weaknesses and strengths. Diversity in the means for conveying knowledge is a good thing. The institutional geolibrary must maintain a complex multidimensional web of mixed media, knowledge sources, collections, and services (Crawford and Gorman, 1995, p. 78). The expectation is that this will be accomplished by merging and embedding geolibrary technological advancements into the existing library information infrastructure of the nation.

DATA, INFORMATION, AND KNOWLEDGE

Geolibraries should play a key role in providing access to carriers of geographic knowledge. In addition, some geolibraries also will want to focus on providing access to the ability to process geographic data. In one sense, all a distributed geolibrary need consist of is a good gazetteer in which users can look up information based on location. The "look up" might be accomplished by drawing a box around an area on a computer screen or by indicating a name

of a place or specifying other information contained in the metadata for a particular item of geoinformation. This would allow a user to find out whether geoinformation covering an area of concern exists in the geolibrary network. If databases exist, the system returns metadata on them so the user can further assess the nature and utility of the databases. Performing this role is consistent with the traditional role of libraries. In addition, gazetteers are one of the prime examples of library documents that were never very efficient in paper form. Conversion to electronic form makes sense since both searching of the gazetteer and updating are made much easier.

If it is legal to copy the databases located through the electronic gazetteer (e.g., public domain geographic databases) or to "check them out" from the holdings within the distributed geolibrary (e.g., the conditions of lending might be determined by licensing agreements), the distributed geolibrary as an institution should be capable of supporting these functions. That is, direct access to the library's holdings should be provided. Again, this function is parallel to and compatible with the traditional roles of the library as an institution.

The level of services and functions (see Chapter 4) provided by the traditional library can be different in geolibraries. Should the services and functions of distributed geolibraries extend beyond providing users with efficient access to the geoinformation in the library's holdings? Or do the technologies that could be provided by distributed geolibraries extend the services and functions in an attempt to provide answers to complex questions rather than guide users to resources where answers may (or may not) be found?

To place this concern in context, we should first define some terms, although no attempt is made here to add to the extensive literature on the nature of information (see, for example, Buckland, 1991; Losee, 1997). The terms listed in the following order are sometimes used to describe an ascending continuum: data, information, knowledge, understanding, and wisdom. Crawford and Gorman (1995, p. 5) define these terms as follows:

Data	Facts and other raw material that may be processed into useful information.
Information	Data processed and rendered useful.
Knowledge	Information transformed into meaning through action of the human mind, such that it can be recorded and transmitted.
Understanding	Knowledge integrated with a world view and a personal perspective, existing entirely within the human mind.
Wisdom	Understanding made whole and generative within the human mind.

Whereas the substantive content and focus of geographic infrastructure building have focused on data and information (e.g., the NSDI), the substantive content of traditional libraries has focused on collections of knowledge and to a lesser extent collections of information. Traditional libraries collect and catalog primarily knowledge works for good reason. The reading and contemplation of works of knowledge such as books and journals provide context and convey meaning. Currently, such works are one of the best means by which we are able to acquire understanding. "Works of knowledge" are largely synonymous with "intellectual works" and are thus the primary expressions protected by our intellectual property laws.

Intellectual Property Concerns

The goal of copyright law, and the effect of copyright law in library settings, has been to strike a balance between giving authors sufficient incentive to make their works available on the one hand and supporting the rights of users to use the intellectual works of others for socially constructive purposes on the other. This balance is complex, but the balance in interests supported by our current intellectual property laws has made libraries highly successful and valued social institutions. A similar balance of interests has not yet been achieved in the online world. A background discussion of current intellectual property and copyright issues

possibly related to distributed geolibraries appears at the end of this chapter.

There is a growing collection of geoinformation available online that is in the public domain because no copyright can exist in some databases due to their nature (e.g., those with no creativity or originality in the arrangement of facts). Claims of copyright in some databases have been rescinded by the authors, and the copyright for other works has expired. Additionally, there are the equivalents of online bookstores and online mapstores that sell or license databases to customers. For the commercial products, libraries have explored several licensing arrangements that attempt to bring together commercial interests and public rights interests to arrive at solutions that support the interests of all stakeholders (see, for example, Barker et al., 1995, and Gladney and Lotspiech, 1998).

In the vision of distributed geolibraries, there is a possibility of creating knowledge and making it available through the distributed geolibrary itself; this raises additional concerns about the status of such derivative knowledge from the perspective of rights and intellectual property. Collections of information, by contrast, gain very little protection under copyright law principles.

Finding 6
Developers of distributed geolibraries will need to consider issues related to intellectual property rights. There are significant differences in both the public access library model and the commercial bookstore model that need to be considered in the broader international debates about the nature of electronic information and databases as intellectual property.

Uses of Data, Information, and Knowledge

Suppose a student wishes to know more about Yosemite National Park and has access through a distributed geolibrary to two different types of information: a digital elevation model (DEM), giving the elevations of points spaced 30 m apart across the park; and a landscape description by John Muir. In principle both are descriptions of terrain, but one is a raw database of measurements

in the public domain, and the other is a creative knowledge work. In another example the couple searching for a home in Chapter 1 might access either a database of socioeconomic statistics or a collection of news reports on the changing characters of neighborhoods.

Both types of information are valuable, depending on the circumstances and the skills and requirements of their users. To a distributed geolibrary they both look like collections of bits with footprints, and both are retrievable using the same mechanisms. Traditionally, one might have looked for the raw data in a data archive, such as the EROS Data Center of U.S. Geological Survey, or a Census data center, and for the description in a library. But distributed geolibraries would provide a unified means of access.

In doing so, however, geolibraries raise issues concerning the relative value of the two types of information. To a specialist equipped with sophisticated tools of analysis, the raw data may be more useful than the landscape description and more acceptable as a source of information for scientific understanding. To a student without sophisticated tools, only the description may be of value. Moreover, the work of the scientist may result in the production of new data, to be fed back into the distributed geolibrary for use by others (such as estimates of solar radiation based on topography combined with a suitable numerical model) or the production of knowledge works in the form of journal articles, which might also be added to the distributed geolibrary. In this sense a distributed geolibrary would be much more than a repository of knowledge because it would support the creation of new knowledge by individuals or groups, in addition to the dissemination of existing knowledge. A student might wish to create personal knowledge as a result of investigation and use a geolibrary to share that knowledge with others in the class.

Both forms of information seem indispensable. There are many questions of a geographic nature that cannot be answered by a *right* answer but require careful reflection based on both data and prior knowledge works. Providing new data query, search, and display capabilities and services may be important in *some* distributed

Finding 7
A distributed geolibrary would support collaborative work, such as multidisciplinary research by teams, decision making by groups of stakeholders, and classroom projects by groups of students. It would provide mechanisms for capturing the knowledge that results from such work and making it accessible to others as appropriate.

geolibraries but providing access to digital works of knowledge is likely to be important in *all* distributed geolibraries.

In summary, a distinction needs to be drawn between raw data and knowledge works because they appear different from the perspective of the functions and services of a library and with respect to intellectual property rights. Although the NSDI is concerned primarily with the production and dissemination of raw geospatial data, distributed geolibraries could also provide an effective mechanism for the dissemination of knowledge.

ACCESS

The concept of access in an institutional distributed geolibrary environment has two major aspects. One involves technical efficiency and effectiveness in finding desired geoinformation, determining its appropriateness and authenticity, linking to and acquiring it, and electronically processing it if needed. To enable such access, knowledge works and databases must exist somewhere on the network with sufficient metadata and tools available in the system to allow these tasks to be accomplished.

The second major aspect of access involves the legal and economic ability of the distributed geolibrary as an institution to provide the geoinformation resources desired by its users, either directly or through the network. If access to intellectual works is barred by legal or economic constraints, powerful computational capabilities and user-friendly search software will not be of any use to the user. Legal rights to materials may alone be an insufficient condition, but they are a critical and necessary condition for access. Acquiring legal rights to intellectual works

and databases can cause a financial burden on the distributed geolibrary and the community it serves.

Although distributed geolibrary collections might be anywhere, they must be somewhere. Those institutions or people with the greatest vested interest in ensuring that specific geoinformation is available, maintained, and accessible are logical candidates for providing those specific collections and resources for distributed geolibraries. For instance, local libraries typically focus on the needs of the local community, and therefore local geolibraries would likely be the primary collectors and maintainers of local geographic information of relevance to local culture.

Another major assumption in the traditional library model is that acquisition or access to commercially provided geoinformation will be through institutional, not individual, payments (Hawkins, 1994). Equity is a fundamental principle of library access. To uphold this principle, the community rather than the individual typically pays for the library and its services (Crawford and Gorman, 1995, p. 101). Just as the poorest Americans can freely borrow books from public libraries, so too should they have equitable access to geolibrary services if a community library chooses to provide those services.

An unanswered issue that will be continually debated in the distributed geolibrary vision is that of access to geoinformation in the public domain and traditional library services versus access to geoinformation and services that are only available on a commercial basis. Embedded in this issue are additional issues of public and private rights and intellectual property. These issues—most of which are not unique to distributed geolibraries—are being debated in the broader library community and the digital information arena (see, for example, two 1997 National Research Council reports—*Bits of Power: Issues in Global Access to Scientific Data* and *More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure*).

In pursuing solutions there is a pressing need to develop new legal, economic, and institutional models that support the public goods benefits of traditional libraries while providing sufficient incentives for private individuals, private publishers, and

government publishers to make their geoinformation available through distributed geolibrary settings. The practical benefits and drawbacks of institutional models will need to be thoroughly explored from economic, legal, and organizational perspectives. Prototype models will need to be developed and tested. It is highly likely that the most appropriate incentive models for private-sector firms will vary from the incentive models that might best encourage local, state, and federal agencies to make their databases available through distributed geolibrary environments.

SUMMARY AND ADDITIONAL ISSUES

This chapter discussed many institutional and societal issues that will have to be addressed by distributed geolibraries, especially if they attempt to replicate many of the services and functions of the traditional library. The major issues are summarized in this section, together with other issues that appear important but were not discussed at length at the workshop.

1. How will local needs for and production of geoinformation be accommodated in a library system that has traditionally emphasized access to books and information with a more general than local focus?

2. Libraries are addressing the need for access to electronic information by developing consortia and networks. How will these new institutional arrangements accommodate and affect the development of distributed geolibraries?

3. Traditional libraries play a significant role in archiving and preserving information. Can this role be accommodated by distributed geolibraries?

4. How can distributed geolibraries deal with inequities of access to electronic systems?

5. Will distributed geolibraries have the effect of enhancing more conventional markets for the information they disseminate?

6. Will distributed geolibraries develop as part of existing library arrangements or complement them?

7. Should the services and functions of distributed geolibraries extend beyond providing users with efficient access to geoinformation to include tools to process and analyze information and create new knowledge?

8. How will distributed geolibraries find an appropriate balance between supplying data and supplying knowledge works?

9. How will each custodian site acquire, give access to, and safeguard the geoinformation in its own collections?

10. How will the distributed geolibrary provide instruction and assistance in the use of digital geographic products and databases? Should users from schoolchild to scientist be expected to be their own reference librarians in the distributed geolibraries of the future?

11. As greater numbers of geographic knowledge works and databases are accumulated in the system over time, will it become increasingly difficult to mine useful information from the available flood?

12. How will the records of humankind be conserved in the distributed geolibrary as an institution?

13. While inclusion of traditional works such as maps in library collections caused few personal information privacy concerns in the past, would the geolibrary's provision for access to detailed databases provide a much greater likelihood for personal information privacy intrusions? What are the principles by which distributed geolibraries would operate in order to protect privacy? How may the principles be enforced and what are the means by which safeguards may be provided in distributed environments?

14. If the generation of knowledge works depends on the resources and intellectual contributions of many persons and institutions, how might intellectual property rights in these works be appropriately accounted for and how might each custodian manage such rights?

15. How can distributed geolibraries assure that geographic knowledge works and databases are not rewritten or revised by government, private firms, or others to their own benefit? That is, how may one assure that databases are authentic?

16. What incentives other than or in addition to future economic rewards could be effective in convincing individuals, businesses, universities, government agencies, and others to make their geographic knowledge works and databases available over a distributed geolibrary network?

17. Who should decide what is in and what is out of a distributed geolibrary? Should there be a gatekeeper, modeled on the function of a library subject specialist, or should distributed geolibraries operate on the principle of *caveat emptor*?

Intellectual Property and Copyright Issues: Background and Context for Distributed Geolibraries

Over the past several years there have been discussions nationally and internationally regarding how to best update the copyright and intellectual property laws to reflect the networked environment. Internationally, the World Intellectual Property Organization (WIPO) has taken the lead in initiating debate on these extremely important yet contentious issues. Nationally, the U.S. Congress has considered a host of intellectual property and copyright issues, many of which originated in WIPO forums.

In December 1996, WIPO member delegates from 160 countries met to consider proposed changes to copyright law with a particular focus on the digital environment. Three draft treaties sought to update copyright law concerning works delivered in digital form, to enact protections for performers in and producers of sound recordings, and to enact a new intellectual property regime to protect databases.

At the close of this diplomatic conference, the delegates adopted two new versions of the three draft treaties originally proposed: one relating to copyrighted works in digital form and the second to enact protections for performers in and producers of sound recordings. Consideration of the third treaty regarding database protection was deferred with the recommendation that WIPO convene another session at a later date to consider a schedule for future discussions on database protection. WIPO failed to move forward on the draft treaty for additional database protection for a number of reasons: lack of time to fully consider the draft treaty within each member country prior to the diplomatic conference, lack of time during the conference to adequately address the

draft treaty, and most importantly, deep concerns, indeed opposition, by many delegations to the draft treaty.

Responding to WIPO's actions, members of the U.S. Congress introduced legislation that would implement the WIPO treaties. A series of hearings and ensuing negotiations between concerned stakeholders on a number of issues such as online service provider liability, fair use, preservation, distance education, and more were undertaken throughout 1997 and 1998. On October 28, 1998, President Clinton signed into law the Digital Millennium Copyright Act of 1998.

WIPO's decision to defer action on a draft database treaty did not deter members of the House of Representatives from considering additional intellectual property protections for databases. Rep. Coble (chair, House Subcommittee on Courts and Intellectual Property) introduced H.R. 2652, the Collections of Information Antipiracy Act.This legislation addresses several concerns of certain parts of the information industry, in particular, legal publishers such as Reed-Elsevier and Thompson. They were concerned with the 1991 Supreme Court decision, Feist v. Rural Telephone, which held that comprehensive collections of facts arranged in conventional formats were not protected under copyright and could not constitutionally be protected under copyright. The decision rejected the notion that a compiler's "sweat of the brow" could ever substitute for the "original authorship" that the statute and the constitutional copyright clause require as the condition of copyrightability.

In addition, some members of the information industry were concerned with a 1996 European Union directive on the legal protection of databases. This directive calls for each member nation to implement a database law by the end of 1997. The directive includes the notion that databases created in non-EC countries will not be granted legal protection; thus, a fear of lack of reciprocity is also prompting segments of the industry to advocate new protections.

During two hearings on H.R. 2652 in the House of Representatives, widespread opposition to the proposal surfaced—from the library community, segments of the commercial sector, the scientific and research communities, the education community, and more. Some of the concerns include the following:

- Provisions in the bill would prohibit a transformative use of information—reuse of information to create a new type of product or information resource.

- The exceptions for scientific and educational use are circular and ineffective, and because the legislation is outside the scope of copyright fair use, related library and education exemptions would not apply.
- Overall the bill would fundamentally threaten the basic paradigm of data exchange by providing unprecedented new legal protection for information.
- Provisions in H.R. 2652 would likely increase the costs of research significantly, as scientists and researchers would have to pay for data they now receive for minimal cost.
- Certain provisions would prevent the creation of "value-added" databases by substantially increasing the cost of the information included in the databases. As a consequence, the elimination of competition from value-added publishers would reduce the incentive for established up-stream publishers to innovate and contain prices.

In a letter to Sen. Hatch (chair, Senate Committee on the Judiciary), the Presidents of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine expressed "deep concerns about the proposed changes to intellectual property law" and noted that the legislation "would grant owners of information unprecedented rights in the control of digital information while severely restricting the rights of scientists and engineers—and everyone else—to access and use that information." Moreover, the anticompetitive nature of H.R. 2652 "may have other negative economic impacts on our information economy by raising prices for data consumers, by stifling important activities of commercial users who add value to existing data, and encouraging the unproductive independent recompilation of the same or similar data."

Other significant concerns were noted by the U.S. Department of Justice, Office of Legal Counsel, which raised serious questions regarding the constitutional basis of H.R. 2652. The Federal Trade Commission noted serious reservations with the legislation, commenting that certain provisions could have "deleterious effects on competition and innovation." Finally, the U.S. Department of Commerce speaking to the concerns of the Administration stated that the legislation as drafted could "increase transaction costs in data use, and ... that legislation not create inappropriate opportunities of incentive to 'capture' government information or government-funded data with relatively small investments in maintenance, organization, or supplemental data."

Although H.R. 2652 was passed by the House of Representatives, it was not considered by the Senate. Members of the House and Senate judiciary committees have commented that legislation that increases intellectual property protection for databases will be a priority in the 106th session of Congress.

A common theme throughout the copyright and intellectual property debates in the United States has been the importance of focusing on appropriate public policy choices for the United States, even though this may conflict with the need for harmonization with other countries' intellectual and copyright laws. According to this argument, the pressure from the European Union directive on databases, for example, should not dictate U.S. information policies with regard to the need for additional protection for databases. Given that the United States is the leader in the information industry, there is an appreciation that legislating in this arena could have significant economic consequences if not done correctly.

4
Services and Functions

LIBRARY SERVICES

Digital library developments are redefining the nature of the library, its services, and its limitations. The traditional library focuses on making it easy for the user to identify, find, browse, and retrieve the contents of a book or journal, but its responsibilities end when the item is in the user's hands. Although the contents of books and journals are essentially immutable, in a digital library the information provided is digital and readily manipulated. Many libraries today have holdings of geoinformation, or provide the means to obtain such data from other sites. Some provide geographic information systems and other tools for users who wish to manipulate or analyze data. Users who access data remotely over the Internet now often have a choice between downloading the data to be analyzed by their own software or sending queries and instructions for execution directly on the data's host. When applied to geospatial data, this remote processing is termed the *GIServices* model to distinguish it from the more traditional local processing of the *GISystems* model. For example, sites such as MapQuest (*www.mapquest.com*) use the GIServices model in providing driving instructions based on geospatial data because the analysis is performed by the host and no data are transmitted to the user. On the other hand, sites such as Microsoft's *www.terraserver.com* and various U.S. Geological Survey sites aim to provide data for local processing, following the GISystems model.

The WWW has made everyone a potential publisher and distributor of information, blurring old distinctions between authors, publishers, distributors, and librarians. The important library function of *collection building*, which involves the library staff in making careful decisions about what should or should not appear in the library, has no equivalent on the WWW, where there are no gatekeepers or custodians of quality.

If library information assets can be accessed from anywhere, how will each library determine what to collect or acquire, if anything? In a digital world and barring direct control and restriction on access, a library will be able to leave more general resources to others and to emphasize those information assets that it alone is best qualified to provide. There would be little value, for example, in serving recent issues of a journal if the journal's publisher and other libraries already provide the needed service at no charge. Unique assets might include the products of the parent institution's own research and scholarship, unique information resources donated to the library by bequests, or information on the library's own local region.

In short, the library of the future will be able to make a clear distinction between the services it provides in helping its users find, access, and use information and the information assets that it collects, builds, and maintains itself. Metadata, or data about data, are likely to become much more important, as libraries seek to refine the services they provide by including more and more tools designed to assist in search, evaluation, and use. Just as today's library needs a catalog that tells users where to look in its stacks for given information resources, so tomorrow's digital library will need the tools (cataloging, indexing, abstracting) that help users navigate the vast communications networks and distributed information resources of the future.

This chapter addresses the services and functions of distributed geolibraries against this background of traditional and novel library services. As noted in Chapter 2, the functions and services of a library are often less obvious than and confused with its physical structure. Some, like information abstraction and

collection building, are less obvious than others, like the physical stacks or circulation desk. Some of the services discussed here have long historical antecedents, while others are entirely novel.

DISTRIBUTED GEOLIBRARY SERVICES

A service can be defined generally as a provision of whatever is necessary for installation and maintenance of a machine, organization, or operation. Services for a machine such as a car include those found at a gas station or a mechanics shop. A small consulting organization might provide sales services to its clients, payroll and training services for its employees, and marketing or research services to maintain steady growth.

The services of a distributed geolibrary fall into several categories, including services for search and retrieval of items of particular interest, item description and display services, data-processing services, and services for collection maintenance and growth. These classes of service relate to the four types of activity that go on in any library: (1) looking for specific books or other reference information by author, title, subject, or identifying code; (2) creation of the library catalog; (3) using various library tools to manipulate or interpret information; and (4) taking care of or improving the library collection.

The nature of these services differs dramatically in a distributed geolibrary, however. The ability to manipulate data, and to integrate data from a number of sources, is greatly enhanced because all data are in digital form. While location was handled as one of a number of possible forms of subject in the traditional library, it is the primary basis of search in a distributed geolibrary. The distributed nature of the geolibrary also makes collection building far more challenging because there are no gatekeepers and no one is in charge of the entire collection.

Moreover, a distributed geolibrary would offer something that is not possible in the traditional library, with its traditional form of catalog—the ability to search based on geographic location.

The power of this concept has already mobilized many individuals, groups, and agencies. For example, the Open GIS Consortium (*www.opengis.org*) issued a Request for Proposals in March 1998 on the subject of catalogs for geospatial data, anticipating that by doing so it would help move the community toward the development of interoperable catalog specifications. The consortium includes roughly 150 vendors, integrators, educators, and users, from both public and private sectors.

THE NEED FOR DISTRIBUTED GEOLIBRARY SERVICES

There are three reasons for developing distributed geolibrary services. The first is economic. Traditionally, geospatial data have been distributed in the form of paper maps, disks, and tapes, which are costly to produce, slow and cumbersome to distribute, and difficult to update. To meet the national mandate to make data collected at public expense available to the public, federal agencies are looking for new ways to disseminate data more widely and effectively, primarily via the Internet (Jones, 1997). By utilizing the Internet and network communications, a distributed geolibrary could deliver online information services quickly and economically. Agencies and companies can also sell data and recover income more effectively using the Internet's growing and increasingly reliable tools for electronic commerce. Finally, encryption technologies could provide assurance against unauthorized use and distribution.

The workshop was not an appropriate forum for the development of a comprehensive economic model of geoinformation dissemination or for detailed analysis of the costs and benefits of implementation. These are important issues and could be the focus of a useful and productive research effort. A good starting point would be a recent study by the National Academy of Public Administration (1998), which includes a comprehensive summary of what is known about the economics of geospatial data production and dissemination.

The second reason involves the decentralization of geoinformation management. In a distributed geolibrary there is no need for data to be collected in one place; instead, data can be held by a custodian until needed. Because the Internet provides universal access, it is sufficient that there be a custodian serving a given data set, and with a single server there are no problems maintaining consistency across copies if data must be updated. Ideally, the custodian would also be the person or agency responsible for updating the data and for assuring their accuracy. In practice, however, some mirroring of data may be needed to overcome the effects of network delays and server downtime (Worboys, 1995).

A third reason for a distributed framework for geolibrary services is the demand for access. Public access to geoinformation, particularly by students, can support improvements in national levels of geographic literacy by making it possible for classes to obtain information quickly and easily about any part of the Earth's surface. Ready access to geoinformation about local areas (neighborhood, city, county, region) can help to develop a more informed citizenry and improve opportunities for participation in the democratic process (Adler, 1995; Craig, 1995).

SERVICES AS COLLECTIONS OF FUNCTIONS

Services have been described using broad categories of response to demands. Functions are the actual commands or activities that implement services, and a given function may contribute to more than one service. A function can deliver all or part of a service. Functions that make up car services at the gas station include changing fluids, changing filters, inspecting brakes or tires, and so forth. At the mechanic shop, the service known as a tune-up would be comprised of functions such as changing spark plugs, adjusting engine timing or belt alignment, and so forth.

Various efforts over the past few years have implemented limited functions of a distributed geolibrary. They include two of the projects of the National Science Foundation-National Aeronautics and Space Administration (NASA)-Defense Advanced Research

Projects Agency Digital Library Initiative (at the University of California's Berkeley and Santa Barbara campuses), efforts of the Federal Geographic Data Committee (FGDC) under the rubric of the NSDI, various state and local government projects; dissemination mechanisms developed by suppliers of Earth imagery, and numerous efforts in other countries. Some selected examples of these prototypes are described in Appendix D. Although there are sharp differences in approach and scope, there is now a degree of consensus on the functions that can best deliver the services of a distributed geolibrary.

NECESSARY DISTRIBUTED GEOLIBRARY FUNCTIONS

Necessary functions for search and retrieval include searches by geographical location, searches by geographical place name, and searches by secondary requirements such as subject theme or time. Retrieval functions require a workspace to hold the items, criteria for sorting and ranking items depending on their assessed relevance to the user's needs, a tagging mechanism to select and retrieve specific items, and links to other functions for display and description. The following sections describe these in more detail.

Search by Geographical Location

The basemap provides the image of the Earth on which a user can specify areas of interest. Its level of geographic detail defines the most localized spatial search that is possible. It should include all of the features likely to be relevant to a user wanting to find and define a search area, including major topographic features and place names. The importance of such features will vary between users, as will levels of detail, so it will be necessary to establish protocols that allow use of specialized basemaps for particular purposes. For example, a hydrologist might want the basemap to emphasize hydrological features such as rivers and watersheds,

whereas a climatologist might want to see weather stations and topography.

This function would first display a basemap, allowing users to point at a place to target either a specific point or a footprint. Users would be allowed to zoom to greater detail and to pan across the Earth's surface. Widgets such as the "rubber rectangle" would allow users to specify footprints in a number of ways. There should also be support for "fuzzy" footprints that are not precisely or crisply defined, allowing users to define approximate areas of search.

There are many current examples of sites that support search by geographic location based on standard WWW browser software (e.g., Microsoft's Internet Explorer or Netscape's Navigator). Many (see, for example, the archive of digital orthophoto quadrangles at the Massachusetts Institute of Technology (*ortho.mit.edu*); other examples are listed in Appendix D) present the user with a map divided into tiles; by pointing to a tile the user accesses data for that tile's geographic area. The Alexandria Digital Library project's current prototype (*alexandria.ucsb.edu*) uses a Java application, including rubber rectangles and other tools. These prototypes use projected basemaps and do not yet implement a sense of interacting with the curved surface of the Earth, as suggested by the vision of distributed geolibraries, which would require three-dimensional visualization technologies such as VRML (Virtual Reality Modeling Language). The current Alexandria browser includes the ability to "paint" data onto the basemap; in Vice President Gore's vision of Digital Earth the user is able to "fly" through a full three-dimensional rendering of the Earth's physical environment.

Several suitable sources of data exist for basemaps:

- Digital topographic data, available for the entire land area of the planet at 1:1,000,000 in the Digital Chart of the World, and for smaller areas at larger scales. For the continental United States the

USGS provides digital topographic data at 1:100,000 and for limited areas at 1:24,000.[1]

- Imagery from space, available from the Landsat satellite at 30-m resolution, from the French SPOT satellite at 10-m resolution, from Russian satellites at 2-m resolution, and anticipated in 1999 commercial satellite imagery for selected areas at 1-m resolution.
- Digital elevation data, available for parts of the United States. at 30-m resolution, and for the entire planet at 5-km resolution. Global coverage at 30-m resolution is planned.

The costs of these data vary enormously; those from federal sources are available at the cost of reproduction, but other sources operate on a commercial basis.

Search by Place Name

Gazetteer is a technical term for an index that links place names to locations. As often found associated with published atlases and city maps, gazetteers provide links to map sheets and locations within map sheets. In the context of distributed geolibraries, a gazetteer connects place names to geographic coordinates. This connection allows the user of the distributed geolibrary to define a search area using a place name, instead of by finding the area on a basemap, which may be difficult to many users. The gazetteer may include place names that are not well defined. For use in a geolibrary a gazetteer must include *extents*, or digital representations of each place name's physical boundary. Links between place names allow searches to be expanded or narrowed—they can be *vertical*,

[1]This ratio or *representative fraction* compares the distance between two points on a paper map with the distance between the same pair of points on the surface of the Earth. Digital data created from paper maps by digitizing and scanning are also characterized by this ratio, which also defines the set of features shown on the map and the degree of geometric generalization of those features. In rough terms a database created from a map with a given representative fraction depicts features larger than 0.5 mm across on the map and achieves a similar positional accuracy.

identifying places that include or are included by other places, and also *horizontal*, identifying neighboring places.

Because a gazetteer is an essential building block of the distributed geolibrary and something that can be shared between large numbers of users, its availability is a critical factor in progress toward the vision of distributed geolibraries. At this time no one agency is identified as being responsible for production and maintenance of a common national or global gazetteer. Most gazetteers that exist, such as the USGS Geographic Names Information System (GNIS) or equivalent commercial products, provide in most cases only a central point for each feature, and their coverage of the world's place names is uneven. Progress would be aided by identification of the gazetteer as a fundamental component of the NSDI framework. Progress would also be aided by the development of a standard gazetteer protocol to ensure that users or groups of users who create their own specialized gazetteers could use them to access distributed geolibraries in place of general-purpose gazetteers. Additionally, there are significant problems to be overcome in dealing with varied alphabets, diacritical marks, ambiguities of spelling, place names with indeterminate boundaries, and so forth.

Search by Subject Theme or Time Period

In a physical library the card catalog indexes library holdings by subject domain. An electronic catalog may include a thesaurus, which matches synonyms of search topics, providing associations in a search query, for example, between "slough" and "swamp" and "wetland." For cataloging functions to work, items must be stored in a standard format, following an agreed protocol. Likewise, users must also specify searches in an agreed protocol; this is often accomplished by a query dialogue function, which converts a form-based user search request into whatever protocol is required. The basis of such protocols already exists in standards, e.g., MARC (MAchine Readable Cataloging, see *lcweb.loc.gov/marc/*) and the FGDC's Content Standards for Digital Geospatial Metadata (*www.fgdc.gov*), and in projects such as the Alexandria Digital

Library. The FGDC has also made progress in standardizing conventions for naming geographic features, and similar progress has been made in other countries.

Distributed geolibraries should allow their users to narrow specifications of need by including subjects, dates, and other identifying characteristics, as well as needed level of geographic detail, and imposing them on the search in addition to geographic location. Although location is the primary key in searching a distributed geolibrary, other aspects allow the user to limit the number of items of geoinformation identified in a search to reasonable levels. Distributed geolibraries also should be capable of ranking items identified in a search by their suitability to the user's needs. They also should inform the user of the number of hits, and provide other ways of summarizing them in readily understood ways.

Item Display and Description

These functions include visualization tools and metadata browsing tools. Visualization tools are useful for displaying items retrieved from the archive. Geoinformation data sets are often massive, creating problems for users who may need to browse through many data sets to find one that is suitable for use, given the limited bandwidth of many Internet connections. In such cases it is clearly impossible to examine the full contents of each data set, and some system must be devised to allow users to examine a summary or generalized sketch of the contents that can be retrieved quickly. Display functions also make it possible to create a visual index (the base map and the map browser, described above) for patrons to search the library for information about a particular place.

In general terms, metadata describe the content, quality, condition, and other characteristics of data. The major uses of metadata include (1) managing and maintaining an organization's investment in data, (2) providing information to data catalogs and clearinghouses, (3) providing information to aid data transfer and use, and (4) providing information on the data's history or lineage.

Although the second use is essentially the function performed by the traditional library catalog, it is clear that the functions of metadata in the distributed geolibrary extend well beyond this (FGDC, see *www.fgdc.gov*). Under (3), metadata provide the essential information necessary to allow a data set from some distant archive to be recognized and opened at the user's site. In general, geoinformation data sets are not *interoperable* in this way, especially if the archive and the user have adopted different geographic information systems (GIS). Problems of interoperation between GIS are addressed by Goodchild *et al.* (1998), and much recent work by the GIS industry has gone into improvements in interoperability in GIS, through the efforts of the Open GIS Consortium (*www.opengis.org*). Note, however, that the problems of distributed geolibraries in this area go well beyond those of GIS interoperability because distributed geolibraries are not just limited to geospatial data.

In the context of distributed geographic information services, metadata include information that supports the exchange of processing operations between client and server (Open GIS Consortium, see *www.opengis.org*). To date, little research has reported on formalization of such metadata to describe distributed geographic information services, though Tsou and Buttenfield (1998) showed that they should include two major parts: *system metadata* and *data operation requirements*. The system metadata describe methods and behaviors for system controls and program specifications, whereas data-operation requirements specify the requirements for data input to, and output from, specified operations.

Collection Creation and Maintenance

A range of tools are needed to support the creation and publication of geoinformation. Most new geospatial data are either published in digital form or go through a digital stage during production. But the predigital legacy of geospatial data is largely in the form of paper maps and photographic images, which must be laboriously digitized or scanned to be suitable for distributed geolibraries. Although massive investments have been made in

recent years, by such organizations as the Library of Congress, which has made much of its historical map collection available over the WWW, it is doubtful that the vast majority of the larger legacy residing in scattered collections and archives will ever be digitized because anticipated levels of use of most individual items cannot justify the cost.

The nation currently possesses vast stores of data about or associated with geographic locations but for which no locational footprint is readily available. These stores include large archives of information on health, the economy, social conditions, and demographics, broken down in some cases to very fine levels of geographic detail. Such data could be incorporated into distributed geolibraries, and *place* could provide a very effective search mechanism, particularly when such data need to be integrated with other geoinformation. A coordinated plan is needed to link as much of this information as possible to geographic location. For example, use of census data could be considerably enhanced if the names and extents of its reporting zones (census tracts, counties, metropolitan areas) could be organized in gazetteer form for use in distributed geolibraries.

Effective description of geoinformation can be difficult, and the FGDC's Content Standard for Digital Geospatial Metadata extends to several hundred fields. While federal agencies are mandated to create such metadata and have access to extensive resources, there is often little incentive for a local agency to create metadata for its own holdings. Many agencies have suggested simplifications of the FGDC standard; the Alexandria Digital Library (*alexandria.ucsb.edu*), for example, uses a subset of 35 fields to describe its holdings. Dublin Core is another effort to simplify the description of information using standard fields (*purl.org/dc*).

The WWW makes it possible for virtually anyone to contribute information by creating and maintaining a WWW site. Distributed geolibraries could take great advantage of this potential by making it possible for users to double as providers of information, especially information that is the result of abstraction,

manipulation, interpretation, or synthesis of other information. For example, papers written based on distributed library resources could be contributed back to distributed geolibraries. The distinction being made here between raw data and derived knowledge is discussed at greater length in Chapter 3.

Searching over Distributed Assets

In a traditional library the catalog provides an index to the library's contents. In a distributed geolibrary the contents and the users are distributed, and five options can be identified for the catalog:

1. A *unified catalog* exists in one place and can be searched by users. In this option each custodian of data submits metadata describing each available data set to a central site, where the records are assembled into a searchable database. Each record directs users to the appropriate location of the data set. For geoinformation, which tends to use specialized formats, this option requires the strongest central control and the highest level of cooperation from participating custodians.

2. Each custodian of data assembles metadata describing each data set according to a standard, forming a *distributed catalog*. Users submit requests to a central site, and these are then automatically executed by search agents that examine each custodian's metadata. Performance of this solution degrades as the number of custodians increases.

3. A *collection-level catalog* exists that identifies the general characteristics of each custodian's holdings and uses them to direct searches. For example, searches for data on some part of New York state might be directed to a custodian in Albany known to have a large collection of that state's data. The efficiency of this option depends on how precisely custodians' holdings can be differentiated. In effect it implements the kinds of expert knowledge that allow users to find data on the WWW in the absence of effective cataloging.

4. Use a catalog built by a *search service*. Search services such as AltaVista and Yahoo build catalogs automatically by usin intelligent agents or *web crawlers*, but they do so strictly on the basis of words found in text and are not effective ways of building a catalog for a distributed geolibrary. Nonetheless, it may be possible to build a new generation of specialized agents capable of recognizing geoinformation and extracting its important metadata descriptors. Such agents have been built on a prototype basis in the case of imagery; they successfully recognize the formats of imagery, open them, and compute such indices as shape, texture, and color for use in catalogs.

5. *No catalog* exists. This reflects the situation on the WWW before WWW search services became available (and even today substantial parts of the WWW's resources remain unindexed by search services). Search for geoinformation without a catalog relies on the user's personal knowledge of the WWW's resources. Whereas a user of a research library can assume with some confidence that any research library will contain a copy of a major monograph or a popular journal, the principle of the WWW is almost exactly the opposite: a given item of information is most likely available at only one site. Search under these circumstances can be like looking for the proverbial needle in a haystack, with order 10^7 sites to search. In the case of geoinformation, the likelihood that a given item will be on a server increases with proximity to the item's footprint for several reasons: interest in the item is likely to be higher near or within the footprint; custodians in proximity to the footprint are more likely to have responsibility; and sponsorship of the data's collection and acquisition is more likely closer to the footprint. But the effect is likely to be weak and as such will provide an unreliable strategy for search.

Integration, Analysis, and Manipulation

Unlike books, which exist largely to be read, much geoinformation is raw in nature and is obtained for purposes that include detailed interpretation, analysis, and manipulation. A user requesting a remotely sensed image, for example, might submit it

to extensive operations that include correction for various known distortions, classification, and integration with other data obtained through similar processes. The end result of obtaining a Landsat image from an Internet site might be a statistical assessment of the amount of change that has occurred in an area over the past 10 years, following several months of detailed analysis and manipulation.

Digital libraries differ from their traditional predecessors in the potential to support extensive manipulation of information once it has been retrieved. This manipulation might include:

- statistical correction for known distortions;
- tabulation to obtain statistical summaries;
- rubber sheeting to register geospatial data sets to known locations or to each other;
- format conversions, projection changes, and datum changes;
- use as input to complex environmental models for purposes of calibration or prediction;
- use in complex decision-making processes involving many stakeholders; or
- generalization, classification, interpretation, and other forms of information abstraction.

Finding 8
A distributed geolibrary would allow users to specify a requirement, search across the resources of the Internet for suitable geoinformation, assess the fitness of that information for use, retrieve and integrate it with other information, and perform various forms of manipulation and analysis. A distributed geolibrary would thus integrate the functions of browsing the WWW with those of GIS and related technologies.

Over the past three decades there has been enormous progress in the development and adoption of technologies for manipulating geoinformation, including GIS and image-processing systems. Today, most users of such systems rely heavily on the

ability to obtain input data from Internet resources, despite the lack of effective tools such as those envisioned for distributed geolibraries. Five steps characterize this gathering process:

1. *Specification* of requirements, including coverage area, date, theme, level of detail, and other important characteristics.

2. *Search* over known or likely sources, using a combination of personal knowledge and the limited capabilities of Internet search services.

3. *Assessment* of the fitness for use of possible data sets, by comparing their documented characteristics with the specified requirements.

4. *Retrieval* of suitable data sets.

5. *Opening* of retrieved data sets on the user's system, including necessary changes of format and other steps needed to integrate data effectively.

Many uses of geoinformation involve group activity—multidisciplinary research projects involving several investigators, planning projects involving several stakeholders and decision makers, group classroom projects involving several students. Distributed geolibraries should provide services to support such collaborative work (see Finding 7, Chapter 3).

Many of the activities that could benefit from distributed geolibraries are best carried out away from the office desktop in the field. Emergency relief operations call for decisions that are best made in the presence of the emergency, where the emergency and its context can be observed directly. Access to distributed geolibraries could usefully augment the power of other field-based technologies, including the Global Positioning System, and mobile computing. Wireless connections could be used to search for and download information from distant servers and to upload new information gathered in the field.

Finding 9
Many important applications of distributed geolibraries are best located in the field, using portable systems and wireless communications. Delivery of services to the field is important in emergency management, agriculture, natural resource management, and many other applications.

Assisting Users

Although the demand can never be fully satisfied, libraries provide large amounts of assistance to their users, funded through library budgets. The Internet provides limited assistance, and users of the WWW are very much on their own, forced to rely on the limited assistance of online help, manuals, and other devices. If distributed geolibraries are to function as a more powerful evolution of the library model, effective ways must be found to help users navigate through their complexities and ambiguities. The problem is, if anything, more severe for geoinformation, which has always required a disproportionately high level of human assistance and user expertise.

We have little experience with the problems that are likely to occur when inexperienced users begin to make widespread use of geoinformation. Problems posed by the important metadata variable *level of geographic detail* are discussed in this context by Goodchild and Proctor (1997), who conclude that new metaphors are needed to make it possible for general users to conceptualize their needs. For example, the metaphor of height of viewpoint above the surface of the Earth (move higher for less detail, descend for more detail) can be readily understood and used by children.

Assessment and Feedback

Libraries also employ staff who *listen* to their users, another function that is difficult to replicate in the impersonal digital environment of the Internet. On the other hand, many new and exciting mechanisms for eliciting feedback have been developed on the WWW, and distributed geolibraries would do

well to exploit these. For example, each custodian site might invite comments on its geoinformation from users and make these remarks available to others. Extensive assessment will be needed of the designs of user interfaces, to evaluate whether they achieve the objectives of distributed geolibraries, before they are widely released and adopted. Such designs should evolve through procedures familiar in the field of human-computer interaction, including evaluation studies and interactive refinement.

OPTIONS FOR THE DELIVERY OF DISTRIBUTED GEOLIBRARY SERVICES

Ideally, we see a distributed geolibrary functioning as a single homogeneous entity capable of responding to a single query from a user, just as AltaVista is capable of responding to a query about some combination of key words. In practice, however, a number of configurations are possible, combining aspects of the following extremes:

1. *One-stop shopping.* One server provides a one-stop shopping service, perhaps to a limited user base via an Intranet or to a universal base via the Internet. Either the entire catalog is mounted on the server or a query to the server results in transparent access to a distributed catalog. Similarly, geoinformation resources are served either directly or transparently through automated access to distributed resources. The agency operating the central server also maintains it, develops and enforces standards and protocols, and directs future development. Several servers currently approximate this mode of operation over substantial thematic and geographic domains, including the USGS's EROS Data Center, and NASA's EOSDIS. This option works well in areas where the resources to create geoinformation come from a single source that can also fund dissemination. Problems arise when jurisdictions or thematic areas overlap significantly. For example, are data about the city of Atlanta more likely to be found in a server operated by the city, county, state, or

federal government, or by the United Nations? Are data about soils most likely to be found on a server operated by the U.S. Department of Agriculture or the USGS?

2. *Distributed responsibility.* This option follows the example of the WWW, for which policies are established by volunteer grassroots organizations that recognize need, devise solutions, and make them freely available to the user community. Protocols and standards allow any individual or group to participate in distributed geolibraries, subject to very loosely defined constraints. Whereas this model approximates the WWW, it differs sharply from the mode of operation of the traditional library, with its substantial resources, gatekeepers, and quality control. The function of cataloging on the WWW, for example, which is approximated by the search services, exists because certain companies saw business opportunities in providing a service that was compatible with WWW standards and met an obvious need. Similarly, quality control in distributed geolibraries might be achieved not by a central gatekeeper authority but by independent groups analogous to the Good Housekeeping Institute that assess and certify geoinformation on a for-profit or nonprofit basis.

Finding 10

There are several alternative architectures for distributed geolibraries, including a single enterprise sponsored by a well-resourced agency, analogous to a national library; a network of enterprises with their own sponsors, analogous to a network or federation of libraries; and a loose network held together by shared protocols, analogous to the WWW.

Geolibrary services can be freely combined and used based on application needs. For geolibraries to operate in a distributed (client-server) computing environment, services and functions must operate on a network of servers and clients. The availability of services must take into account server characteristics, such as file sharing and application serving, and whether there is a "thin" or "thick" client. In networking terminology a thick client is

defined as having operations and calculations executed on the client, consistent in this context with the GISystems model. A "thin" client may require that selected functions run on the server, consistent with the GIServices model. Whether the client should be thick or thin will depend on the task and associated performance requirements. For example, it may be appropriate to use thick clients for map display services, allowing the patron to take over the many intuitive decisions of graphic design, layout, and so forth, as well as to accommodate whatever output devices are available.

Still other functions may deliver services best by avoiding transmission of large amounts of repetitive data across a network. For map browsing and place name searches, a geolibrary might use a "hybrid" approach by storing the basemap and gazetteer on the client but leaving the catalog functions on the server. Basemap information is voluminous and not likely to change frequently, so rather than transmit it repeatedly from a server it may be more efficient to store it locally in a specialized hybrid browser. Suitable basemaps include digital topographic maps and also images of the Earth's surface. Additional detail can be provided by digital elevation data, so the basemap provides a close resemblance to the actual surface of the Earth. "The role of client and server components should be dynamic and changeable. The balance of functionality between client services and server components will be a critical issue for the success of ... distributed systems" (Tsou and Buttenfield, 1998).

5
Building Distributed Geolibraries

REQUIREMENTS

Previous sections of this report outline the vision of distributed geolibraries, discuss the problems and issues related to their social and institutional context and define their services and functions. This chapter addresses the process of building distributed geolibraries, the steps that will need to be taken to implement the vision, and related issues. It is impossible to be precise, of course, because of uncertainties surrounding future technologies, because the outcomes of research are in principle impossible to anticipate, and because many issues can only be resolved by constructing and working with prototypes. Given these constraints, this report attempts to address a number of key questions and to find answers where possible:

- What will it take to build distributed geolibraries?
- What economic incentives can be put in place such that stakeholders in all sectors of the community (business, education, government) can and will participate?
- What arrangements need to be put in place in the form of institutions, regulations, standards, protocols, committees, and so forth?
- What research needs to be done to address problems and issues for which no methods or solutions currently exist? How long will this research take?

- What data sets need to be constructed, and what mechanisms might be used?
- What software needs to be written, and who is likely to write it?

At a higher level one might ask how it is possible to know the answers to these questions. Complex software systems and new institutions arise through an iterative process in which the end result may not be apparent until the process has been under way for some time. Creating a vision is part of that process, but the vision may be wrong or unachievable. Large-scale prototypes are sometimes built in part because it is difficult or impossible to know what is possible without such large-scale experimentation. Without building a distributed geolibrary prototype, it may not be possible to identify exactly what it will do successfully and what it will not do. It may be difficult to know at an early stage how much a distributed geolibrary will cost or whether its costs will be exceeded by its benefits.

The Panel' vision of distributed geolibraries views them as a primary distribution mechanism for getting geospatial data and geographic knowledge resources into the hands of all stakeholders. Traditionally, the primary source of geospatial data in the United States, as in many other countries, has been the national mapping agency. Dissemination has been predominantly a one-to-many operation, as a single source provided information to a distributed user base. The vision of the National Spatial Data Infrastructure (NSDI) is very different and reflects an increasing degree of empowerment of individuals and agencies as significant producers of geospatial data. This vision is many-to-many, replacing a single source with a much more complex array. It is also complicated by the fact that the user/producer distinction is no longer as clear. Many users of geospatial data add value and become producers, and many users serve their own networks of clients. Many users of geospatial data are producers of geographic knowledge, which they may want to publish or make available through the mechanism of distributed geolibraries.

The many-to-many paradigm is familiar to librarians, who have traditionally acted as brokers between the publishers and the users of information. Thus, the paradigm shift that is occurring in geospatial data dissemination, in part through a process of technological empowerment, provides a strong reason to look to the library as a metaphor for new dissemination models and suggests that the library is a good place to look for models of distributed geolibraries and for solutions to problems and issues that may arise in building them. On the other hand, the timescale of library operations has been far slower than is normal with digital data dissemination. It may take years for information to pass fully through the complex process of publication and cataloging until it is finally available to the traditional library user. Users of the WWW are accustomed to delays on the order of minutes not years. Thus the library model will be useful only if its customary timescales can be compressed by many orders of magnitude.

The following sections address the needs of distributed geo-libraries in terms of standards and protocols, data sets, georeferencing, cataloging, visualizations, and knowledge creation. Later sections discuss research needs and institutional arrangements. The final section of the chapter discusses the measurement and assessment of progress in building distributed geolibraries.

Standards and Protocols

Geospatial applications are already supported by a large number of standards and protocols, and many more are in various stages of development. The set of particular relevance to distributed geolibraries includes:

- The metadata standard developed by the Federal Geographic Data Committee (FGDC) and known as the Content Standards for Digital Geospatial Metadata (*http://www.fgdc.gov*). This standard allows catalogs of geospatial data sets to be constructed using well-defined content. It is elaborate, and substantial effort is needed to achieve compliance. A very similar general metadata standard is in

the International Organization for Standardization (ISO) review process under the ISO Technical Committee 211 (ISO-TC211).

- General file format standards for geospatial data. These include standards mandated under FIPS 173 and known as the Spatial Data Transfer Standard (SDTS), the scientific data standards HDF and netCDF, the imagery standards TIFF and GeoTIFF, the military standard DIGEST, and many more.
- Interoperability specifications. The Open GIS Consortium (*www.opengis.org*) is developing a wide range of specifications for geospatial objects to support interoperation and is strongly supported by the GIS software industry.

Other standards of relevance to distributed geolibraries include those under discussion on intellectual property rights in digital data, standards of geospatial data quality, definitions of geographic feature types, and general mapping standards. They are being developed through a multitude of standards organizations, including, for example, the ISO, the American National Standards Institute (ANSI), the FGDC, and the International Cartographic Association.

The Internet and the WWW are built on a series of standards and protocols that have been widely accepted not because of any compulsion or mandate but because they clearly work and enable interesting applications. They include TCP/IP and HTTP. In the coming years it is likely that these standards will be extended repeatedly, and it appears that the architecture of the Next-Generation Internet will be significantly enhanced. Although none of these developments have been driven or are likely to be driven by the special needs of distributed geolibraries, as in the past we can expect them to be exploited in whatever ways are interesting, valuable, and appropriate.

Finding 11

New technological initiatives such as the Next-Generation Internet and Internet II are likely to provide extensions to Internet and WWW protocols and orders of magnitude increases in bandwidth. Many of these developments are expected to be relevant to distributed geolibraries.

Data Sets

Libraries assist their users in many ways; some of the most important are the mechanisms of abstraction employed to help users find relevant information. The process of cataloging is assisted by a number of data sets known as *authorities* that provide essential indices and lists.

In distributed geolibraries an essential authority is the gazetteer. A distributed geolibrary's gazetteer will differ in several key respects from the traditional version found in the back pages of atlases:

- Support for *extents*, defined as the bounding coordinates of place-names. Traditional gazetteers, and their digital equivalents such as the Geographic Names Information System provide only point references for most features. In contrast to point locations, extents are needed to resolve the relevant discrepancies between the given footprint of an asset and the footprint of a user query. Because there is only marginal value in a highly precise footprint (since adding additional precision to a boundary's location will only marginally increase the effectiveness of a search), it may be sufficient to provide only bounding coordinates (e.g., minimum and maximum latitude and longitude).
- *Extensibility*, defined as the ability of a user to insert additional place names of interest into a local copy of a standard authority gazetteer.
- *Specialization*, defined as the ability of a user to define gazetteers for special applications. Many application domains have their own equivalents of recognized place names. Hydrologists use standard ways of indexing watersheds, for example, and remote sensing specialists use standard numbering systems for the images derived from satellites such as Landsat. Translations from these systems to standard coordinates will be important data sets in support of the functions of distributed geolibraries.
- Support for *fuzziness*. Traditional gazetteers literally provide authority only for officially recognized place names. While the

footprint of a city name may vary depending on context and usage, the official footprint is most often defined by the city limits. Users of distributed geolibraries will want to be able to search based on place names that are not officially recognized but nevertheless in common usage, such as "downtown."

Finding 12
A comprehensive gazetteer, linking named places and geographic locations, would be an essential component of a distributed geolibrary. A national gazetteer would be a valuable addition to the framework data sets of the NSDI. These framework data sets are being coordinated by the FGDC, which also has the responsibility for associated standards and protocols. Production and maintenance of the national gazetteer could be through the National Mapping Division of the U.S. Geological Survey (USGS) in collaboration with other agencies and could be an extension of the USGS's Geographic Names Information System.

Another type of authority used by libraries is the thesaurus. In the geoinformation case, various kinds of authorities would be useful: lists of standard feature types, standard data themes, standard attribute definitions. For example, it would be useful if the meaning of *vegetation* and associated terms could be standardized, and much effort by the FGDC has been devoted over the past few years toward this end. In a world in which everyone can be a data producer, it is no longer possible to rely solely on the federal government to define essential mapping terms.

At the same time it is important that distributed geolibraries reflect the contemporary social norms of their users. The very term *authority* suggests a command-and-control philosophy that may be orthogonal to the prevailing culture of the Internet and the WWW, which is dominated by individual empowerment and voluntary consensus. An authority for a distributed geolibrary is clearly something different from a traditional library authority, and digital technology must be used to serve different ends. Instead of a single authority created by a central agency and enforced *top-down* on the community through regulation, mandate, or incentive, digital

technology should be used to support translation and interoperability between a variety of different meanings and interpretations in a *bottom-up* process that accommodates diverse communities and groups and their associated terminologies. If the term *downtown* means something different to user A than to user B, distributed geolibraries should use the power of digital technology to make the two meanings interoperable, rather than to support the imposition of a single interpretation on all users.

Georeferencing

The system of latitude and longitude has been subject to international standards since the late nineteenth century. However, the definitions of latitude and elevation are dependent on the mathematical function used to approximate the shape of the Earth, and many such functions are in use. Thus, latitude is not fully interoperable, and two points near each other on the Earth and measured from opposite sides of certain international boundaries do not converge perfectly. Additional complications occur in the use of other world coordinate systems, such as UTM (Universal Transverse Mercator coordinate system) and between the U.S. State Plane coordinate systems. If distributed geolibraries are to be useful to people who do not understand the complexities of geodetic datums and cartographic projections, it will be necessary for systems to be developed that are capable of hiding such details or making them fully transparent to the user. Thus, a user ought to be able to access data sets in different projections and based on different datums and expect the system to handle the differences automatically. *Such transparency is not yet available in standard geospatial software products and data sets, and its feasibility has not been demonstrated.*

Other general ways of referencing the surface of the Earth are gaining popularity because of interest in global environmental change and other processes that operate at the global level. These include standard hierarchical grids such as QTM (Dutton, 1984) and the sampling grids used by the EMAP program (White *et al.*, 1992).

Such hierarchical systems may be important internally as indexing schemes for distributed geolibraries (Goodchild and Yang, 1992).

Cataloging

Reference was made earlier to the need to compress the traditional timescales of the library world. Nowhere is this more important than in cataloging, which serves the critical function of abstracting the information users need to find, examine, assess, and retrieve data. In effect, metadata are the key to the many-to-many structure that allows many users to search across many potential suppliers, and its timely creation will be crucial if distributed geolibraries are to function. Unfortunately, the process of metadata creation for digital geospatial data can be as lengthy and labor intensive as its traditional equivalent. The task of creating a full metadata record for a geospatial data set using the FGDC metadata standard can be much greater than the task of cataloging a simple book. The geospatial data community appears to have accepted the notion that metadata creation is largely the responsibility of the producer, whereas the prevailing notion in the library community is that cataloging is the responsibility of the librarian. This reflects a distinct difference in philosophy, since the library practice is based on the notion that the librarian may be *more skilled* in abstracting information on behalf of the user than is the producer of the information.

If time is of the essence in the digital world of the Internet, it makes good sense to try to replace the labor-intensive cataloging process with automated methods. The Internet world's solution to this problem has been the WWW search service, exemplified by AltaVista, Yahoo, and Excite. To be successful, a search service designed to help the user of distributed geolibraries find geospatial data and geographic knowledge would have to place heaviest emphasis on the determination of an information object's geographic footprint, either by detecting or inferring coordinates or by identifying an appropriate place name, to be converted to coordinates using a gazetteer. Such tools would perform the functions of abstracting and metadata creation automatically. *Such*

automated discovery, indexing, and abstracting tools do not yet exist and will require extensive research and development. Three models that provide alternatives to the search service are described in Chapter 4. They are technically much simpler, but require practices that appear to be incompatible or only partially compatible with the culture of the Internet.

Visualization

One of the most powerful advantages of the concept of distributed geolibraries is the ability for the user to interact with a representation of the surface of the Earth. Information *about* the Earth's surface is naturally conceptualized as belonging to the surface, and globes, which are actual scaled representations of the Earth, provide a familiar and easily understood information source. The notion of doing the same in the digital world, of presenting information as if it were actually located on the surface of the globe, is termed the *Digital Earth* metaphor, and lies behind the idea described earlier in Chapter 2.

Some types of geoinformation illustrate close approximations to actual appearance and can be rendered by draping onto a curved surface. These include optical imagery and false-color imagery, where colors are used to render information that corresponds to some other possibly invisible part of the spectrum.

Other information in distributed geolibraries is not rendered so easily. How, for example, would one portray economic information such as average household income using the Digital Earth metaphor? In some cases there may be clever ways of making visible what is normally invisible; in other cases it may be necessary to represent the presence of information using symbols that exploit some other metaphor, such as books or library shelves. *This is a novel area with no obvious guideposts, and research will be needed to determine how best to make the user of distributed geolibraries aware of the existence of information and of its important characteristics.* In particular, we know almost nothing about how to render dynamic geospatial data or how to indicate

availability, yet we anticipate that such data will be increasingly available to the users of distributed geolibraries.

Knowledge Construction

Users of distributed geolibraries will need tools for analysis, modeling, simulation, decision making, and the creation of new geographic knowledge. An important component will be the workspace in which the user can process data using many of the functions found in today's GIS, along with other functions such as those described earlier in Chapter 4. Given the massive investment in GIS, the easiest way to achieve this will be through collaboration between the builders of distributed geolibraries and the developers and vendors of GIS software. Compatibility and interoperability between GIS products and distributed geolibraries will be needed. For example, the metadata used to discover, assess, and retrieve data should be processed and updated by the GIS as data are manipulated and used to create new data sets. Metadata should be generated automatically when new knowledge is created by analysis and modeling. *Current software products are generally incapable of these functions, and much research remains to be done to make them generally available.*

RESEARCH NEEDS

Many of the topics discussed in this report fall under the heading of "things we do not yet know how to do." In some cases, such as the building of a distributed geolibrary itself, there may be no obviously missing piece of theory or understanding; rather, it may be that we have not yet tried and that given sufficient resources the necessary knowledge will be available. But other items require more focused research. Among them are the following:

- *Scalability.* We have no experience with building and operating data-handling systems on the massive scales envisioned here.

- *Interface design*. Most information technologies are designed for skilled users. Distributed geolibraries will be used by everyone, over a wide range of levels of cognitive understanding, and will require new methods of interface design that embody sound principles, some of which have yet to be discovered.
- *Merging data*. We have very little experience with the massive redundancy anticipated in distributed geolibraries, where many sources of the same data will be available. We do not have techniques for merging data from different sources, across different scales and levels of accuracy, or across different data models or ontologies, or for combining or conflating the desirable properties of sources. Distributed geolibraries will be one of a growing number of applications that depend on the ability to register multiple data sets quickly and easily and to remove obvious discrepancies.

Finding 13
The success of a distributed geolibrary will be largely dependent on the ability to integrate information available about a place. That ability is severely impeded today by differences in formats and standards, access mechanisms, and organizational structures. Removal of impediments to integration should become a high priority of government agencies that provide geospatial data.

- *Indexing*. Our methods of indexing data have been developed for the flat two-dimensional world of maps and images. Distributed geolibraries will require comprehensive approaches to indexing that are capable of supporting "drilling down" over a wide range of scales.
- *Visualization*. While techniques for visualizing static two-dimensional data are well understood, particularly in cartography, we do not have the same level of understanding of appropriate ways to visualize data on the curved surface of the Earth, especially when the data are time dependent. Much more research is needed into appropriate metaphors, techniques, and user responses before these will be as easy as traditional cartographic visualization.

Finding 14

Significant research problems will have to be solved to enable the vision of distributed geolibraries. Research is needed on indexing, visualization, scaling, automated search and abstracting, and data conflation. Research on these issues targeted to improve access to integrated geoinformation might be pursued by the National Science Foundation and other agencies sponsoring basic science, as well as by the National Mapping Division of the USGS, and the National Imagery and Mapping Agency.

Many mechanisms and programs already exist to move this research agenda forward. Examples include the following:

- *The Digital Library Initiative.* Funded first in 1994 by NSF, NASA, and DARPA, this program was recently reannounced (*www.nsf.gov/pubs/nsf9863/nsf9863.html*), and is expected to fund research through 2003. Among the six projects funded by the first round, those at the University of California's Berkeley (*elib.cs.berkeley.edu*) and Santa Barbara (*alexandria.ucsb.edu*) campuses are particularly relevant to distributed geolibraries.
- *Digital Earth.* As discussed in Chapter 2, Vice-President Gore described a vision of Digital Earth that bears substantial resemblance to distributed geolibraries. In the next few years this vision may develop into a substantial funded research program.
- *Digital Government.* NSF recently announced research opportunities in a new program to build stronger ties between the research community in computer and information science and engineering and various government departments with very significant investments in systems and data integration (NSF Program Announcement 98-121). This program may be a suitable vehicle for promoting the research needed to support distributed geolibraries.
- *Knowledge and Distributed Intelligence (KDI).* NSF's KDI program announcement (NSF Program Announcement 98-55) has strong relevance to the vision and issues of distributed geolibraries.

• The August 1998 Interim Report of the President's Information Technology Advisory Committee (*www.ccic.gov/ac*) called for substantial increases in federal information technology research and development and for a series of virtual *expeditions* in specific areas. An effort in distributed geolibraries seems to fit the intent of the report well.

In addition to these formal mechanisms, significant research and development activities are under way in the private sector among vendors of GIS software and among defense and intelligence contractors that can be expected to push in the direction of distributed geolibraries over the next few years. For example, the vendors of new commercial space imagery could use systems like distributed geolibraries for the dissemination of their data products to the broad user community. The FGDC is also a potential source of research initiatives in this area, given its relevance to the future dissemination mechanisms of the NSDI.

Many of the research needs identified here are basic in nature, and it may be many years before solutions can be found. On the other hand some issues such as the need for better methods of data integration, are so widely recognized, technical in nature, and strongly motivated that significant progress can be expected in a comparatively short period.

INSTITUTIONAL NEEDS

Although elements of a distributed geolibrary already exist in the form of prototype clearinghouses and other projects, it is easy to lose sight of the broader concept and the degree to which it represents a radical departure from current and past practices as reflected in our institutions and their accepted functions. More specifically:

• Traditional production and dissemination of geoinformation have been *centralized*, as functions of the upper levels of government. These arrangements made good sense in the past, but the empower-

ment that has occurred as a result of the almost universal adoption of information technologies, especially geographic information technologies, over the past two decades has called them into question. Yet such institutions as the national mapping agencies still reflect this legacy. *The vision of distributed geolibraries represents a broadly based restructuring of past institutional arrangements for the dissemination of geospatial data and one that is much more bottom-up, decentralized, and voluntary.* The institutional arrangements of the WWW provide an excellent model.

- The implications of distributed geolibraries for intellectual property rights, the library as an institution, and the economics of information use are discussed at length in Chapter 3.

- Traditional production and dissemination practices for geoinformation have emphasized the *horizontal* integration of information at the expense of *vertical* integration. Today it is much easier to obtain and make use of the *same type of data for different areas* than it is to obtain and make use of *different types of data for the same area.* A distributed geolibrary would prioritize vertical integration to obtain responses to such queries as "What have you got about *there*?" Producers and distributors of geospatial data could make it much easier to integrate different types of data. The USGS, for example, could make it easier to obtain digital elevation data, digital topographic data, and digital orthophoto data for the same area. Today that ability is severely impeded by differences in formats and standards, access mechanisms, and organizational structures, as well as in the basic geometric and positional problems associated with varying accuracy and varying definitions of shorelines and other features.

Finding 15

While traditional production of geospatial data has been relatively centralized, the vision of distributed geolibraries represents a broadly based restructuring of past institutional arrangements for the dissemination of geospatial data and one that is much more bottom-up, decentralized, and voluntary.

Some of these issues are specific to geoinformation and geospatial data, but others are generally applicable to the emerging information society, which is being driven by technological change and by the desire for greater access to information. Lopez and Larsgaard (1998) discuss this relationship between the needs of the geospatial data and the broader institutional setting of the evolving digital library. That relationship is complex, and it is clear that distributed geolibraries are part of a larger vision of the digital library of the future. But the central role they give to searches based on location makes them clearly distinct, as do the research problems identified in the previous section. The development of distributed geolibraries will require a unique set of partnerships between developers of information technologies, geographic information scientists, application domain specialists, and user communities. It is unlikely, therefore, that the vision of distributed geolibraries will be realized through broadly based efforts to research and develop digital libraries in general; instead, efforts are needed that are directed specifically at distributed geolibraries and geoinformation. Funding and coordination are needed to develop prototypes, stimulate basic research, and build partnerships that specifically address the vision of distributed geolibraries.

MEASURING PROGRESS

The workshop convened by the Mapping Science Committee (see preface) was designed to help identify a vision of distributed geolibraries and the steps needed to realize that vision. An important element of building distributed geolibraries is, therefore, the measurement of progress: how will we know how much progress has been made and how much remains to be done? In this section we offer some possible bases for measurement.

- *Query-based.* If the objective of distributed geolibraries can be expressed in the ability to issue the query "What information is available about *there*?", a simple measure of progress can be based on the amount of information available to a user of the WWW in

response to queries of that nature. Some of the sites listed in Appendix D can already respond to that type of query. A simple measure would be complicated by the various conditions under which information is available, such as cost, intellectual property restrictions, and quality.

- *Analysis-based.* Rather than base progress on the availability of data, a more sensitive and powerful measure might be one based on the ability of the user to obtain services that involve analysis. Information that involves processing in its creation from raw data, and information that represents knowledge, can be of more value that the raw data themselves. If distributed geolibraries are to involve a vision of services rather than simple data supply, measures based on the complexity of analysis will be important indicators of progress.
- *Cost-based.* One way to assess a traditional library is on the basis of cost: How much does it save its users to have access to resources such as books or databases via libraries in lieu of the user purchasing them? If economics is the real driver of the library system, the same argument can be made about distributed geolibraries: specifically, how much is saved when data are shared rather than re-created in multiple archives?
- *Abstraction-based.* Another view of the traditional library is that it is a successful abstraction mechanism, allowing its users to find and retrieve information objects (books) without direct knowledge of their contents, through the mechanisms used by the library to abstract and catalog. One might measure the progress of a distributed geolibrary on this basis, by developing indicators of the amount of work required on the part of the user to find a given item of information. The library has clearly failed if this can be done only by inspecting the contents of every information object in the library.

In addition, progress toward the vision of distributed geolibraries could be measured through the volume of accumulated research results, the sophistication of prototypes, and the lessons learned from each.

6
Conclusions

REVISITING THE RATIONALE FOR DISTRIBUTED GEOLIBRARIES

Chapter 1 presents a limited set of examples for which the ability to access information by place from distributed resources would be useful. In the first example, a truck accident has caused the potential for major environmental disaster and possibly loss of life. The accident's impact is directly dependent on the ability of those responding to gather the necessary information on which an effective response strategy can be based. Knowing exactly where the accident occurred can reduce the time taken to make the first response. Knowing what is likely to happen to the spilled liquids or gases can reduce their impact, lead to more rapid cleanup, and avert many possible costly outcomes.

Actual benefits of improved access to information in such circumstances are extremely difficult to estimate. Many of them are intangible and thus difficult to express in dollar terms. Outcomes of such events vary enormously in severity, yet the difference between a life lost and a life saved is immense. In the case of the 1995 Oklahoma City bombing, for example, it has been suggested that the use of a computer-based model of the Murrah federal building, along with simulations of how the explosion modified the structure and of where the occupants were likely to be found, shortened the total duration of the rescue effort by several days and significantly increased the probability that victims would be found alive.

In the other examples in Chapter 1, the value of improved access to geoinformation lies in the intangible benefits of a better-informed citizenry and of improved access by stakeholders to the information resources of governments and other agencies. In these examples, *place* provides by far the most effective means of searching for information when the issue is localized to a neighborhood, city, or region and when it spans many different themes, disciplines, and areas of responsibility. Location is the *only* way to link information from diverse themes in such circumstances, and our current inability to do that is a major impediment to informed debate on many of the issues that concern society.

If a distributed geolibrary in some form is not developed, a major opportunity made possible by recent developments in information technology will be lost. With a geolibrary the time needed to respond to emergencies could be reduced, as those responsible for dealing with emergencies would have vastly improved means to assemble needed information. And with distributed geolibraries the average citizen and stakeholder will have a greater opportunity to be better informed about many local and regional issues.

DISTRIBUTED GEOLIBRARIES IN CONTEXT

Chapter 2 describes a physical geolibrary as a building containing a large globe with which users would specify their areas of interest; in response, the library would provide all of the information relevant to that area. The concept was presented as a thought experiment, since clearly such a physical geolibrary could not be built. However, the concept suggests two questions that should be addressed: (1) To what extent is the geolibrary an extension of the traditional library with its card catalog and search mechanism based on author, title, and subject? (2) How will the geolibrary complement traditional libraries?

The workshop and this report have focused almost exclusively on queries defined primarily by location, arguing that

for many reasons such queries have been difficult to handle in the traditional library and that the kinds of materials best found through such queries are consequently less likely to be found in the traditional library. Thus, the geolibrary is distinguished by both a distinct search mechanism and a somewhat distinct collection.

It is important not to give exclusive emphasis to place-based searches, even in the case of geospatial data. Consider, for example, the following query "Do you have a picture of a hurricane?" Queries of this form are common in education, for example, or the news media, and although they require geospatial data in response, such as an image from space, the data's footprint on the Earth's surface is actually irrelevant to the search.

The Panel suggests that the geolibrary is complementary to the traditional library in the sense that it adds a new search mechanism to the traditional one. By adding place-based search to searches based on author, title, and subject, the distributed geolibrary allows users with needs defined by place to search the distributed archive of the WWW in new ways. In turn it encourages producers and custodians of geoinformation to make their information assets accessible through the WWW. Whether a distributed geolibrary evolves into a distinct set of software, protocols, and institutions or whether it becomes fully integrated into the distributed digital library of the future remains to be seen.

The ability to search by place should provide a strong stimulus to the producers and custodians of geoinformation to add specifications of footprints and to make use of metadata formats that include such information, including the FGDC's Content Standards for Digital Geospatial Metadata (*www.fgdc.gov*) and suitably extended versions of the Dublin Core (*purl.org/dc*). Government agencies could take a lead in this direction by developing a coordinated plan to link as much information as possible to geographic footprints. This is already under way in those agencies represented on the FGDC, since such agencies are mandated to produce metadata according to the FGDC standard for all of their geospatial products. But the United States possesses vast archives of information that could be incorporated in a

distributed geolibrary collection and made accessible to place-based search if it could be linked to a footprint. Linking much of this information to geographic location—in other words, to transform it to geoinformation—would be valuable within a geolibrary context.

Several programs discussed in Chapter 5 might provide support for the development of the distributed geolibrary, although none is targeted to the specific research problems associated with place-based information resources. Funding will be needed to stimulate the development of prototypes, support research, and build partnerships directed specifically at distributed geolibraries, so that the vision outlined in this report can become a reality and the problems of data access identified at the outset in Chapter 1 can be addressed effectively.

References

Adler, P. S. (1995). Geographic information systems and academic libraries: An introduction, *The Journal of Academic Librarianship* (Special Issue) *21*(4): 233-235.

Barker, J., et al. (1995). Rights Manager System: Permissions Manager Subsystem (Version 2 draft), Library Collections Services, Case Western Reserve University.

Borgman, C. L., M. J. Bates, M. V. Cloonan, E. N. Efthimiadis, A. Gilliland-Swetland, Y. Kafai, G. L. Leazer, and A. Maddox (1996). *Social Aspects of Digital Libraries*, Final report to the National Science Foundation; Computer, Information Science, and Engineering Directorate; Division of Information, Robotics, and Intelligent Systems; Information Technology and Organizations Program.
http://www.gslis.ucla.edu/dl/

Buckland, M. K. (1991). Information as thing, *Journal of the American Society for Information Science 42*(5), 351-360.

Corporation for National Research Initiatives (1998). *NSF SMETE-Lib Study: Report of the SMETE Library Workshop.* Washington, DC.
http://www.dlib.org/smete/public/report.html

Craig, W. J. (1995). Public data access: Another side of GIS data sharing, in *Sharing Geographic Information,* H. J. Onsrud and G. Rushton, eds., Center for Urban Policy Research, Rutgers University, New Brunswick, N.J., pp. 107-118.

Crawford, W., and M. Gorman (1995). *Future Libraries: Dreams, Madness and Reality*, American Library Association, Chicago.

Dutton, G. H. (1984). Geodesic modelling of planetary relief. *Cartographica 21*(2,3), 188-207.

Gayon, T. (1998). Local Geographic Information and Technology in Colorado Public Libraries, Master's thesis, Department of Geography, University of Colorado, Boulder.

Gladney, H. M., and J. B. Lotspiech (1998). Safeguarding digital library contents and users: storing, sending, showing, and honoring usage terms and conditions, *D-Lib Magazine*, May 1998.
http://www.dlib.org/

Goodchild, M. F., and J. D. Proctor (1997). Scale in a digital geographic world. *Geographical and Environmental Modelling 1*(1), 5-23.

Goodchild, M. F., and S. Yang (1992). A hierarchical spatial data structure for global geographic information systems, *CVGIP: Graphical Models and Image Processing 54*(1), 31-44.

Goodchild, M. F., M. J. Egenhofer, R. Fegeas, and C. A. Kottman, eds. (1998). *Interoperating Geographic Information Systems*, Kluwer, New York.

Hawkins, B. L. (1994). Creating the library of the future: Incrementalism won't get us there! *The Serials Librarian 24* (3-4), 17-47.

Jones, J. (1997). Federal GIS projects decentralize. *GIS World 10*(8), 46-51.

Lopez, X. R., and M. Larsgaard (1998). Towards a California geospatial digital library: a strategy for networked knowledge. *Cartography and Geographic Information Systems 25*(3), 133-141.

Losee, R. M. (1997). A discipline independent definition of information, *Journal of the American Society for Information Science 48*(3), 243-269.

Moeller, J. (1998). The National Spatial Data Infrastructure: An Overview, *Office of Intergovernmental Affairs Newsletter 5*, U.S. General Services Administration, Washington, D.C., pp. 12-13.
http://policyworks.gov/intergov

National Academy of Public Administration (1998). *Geographic Information for the 21st Century: Building a Strategy for the Nation*, Washington, D.C., 358 pp.

National Research Council (1993). *Toward a Coordinated Spatial Data Infrastructure for the Nation.* Mapping Science

Committee, National Academy Press, Washington, D.C., 171 pp.
http://pompeii.nap.edu/books/0309048990/html/index.html

National Research Council (1997). *Bits of Power: Issues in Global Access to Scientific Data*, Committee on Issues of Transborder Flow of Scientific Data, National Academy Press, Washington, DC, 235 pp.
http://www.nap.edu/readingroom/books/BitsOfPower/

National Research Council (1997). *More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure*, Computer Science and Telecommunications Board, National Academy Press, Washington, D.C.
http://www.nap.edu/readingroom/books/screen/

National Research Council (1998). *Developing a Digital National Library for Undergraduate Science, Mathematics, Engineering, and Technology Education: Report of a Workshop*, Center for Science, Mathematics, and Engineering Education National Academy Press, Washington, D.C.
http://www.nap.edu/readingroom/books/dlibrary

Newcomb, S. R., N. A. Kipp, and V. T. Newcomb (1991). The Hytime hypermedia/time-based document structuring language, *Communications of the ACM 34*(11), 67-83.

Perritt, H. H. (1996). *Law and the Information Superhighway.* John Wiley & Sons, Somerset, N.J:

Rhind, D. W. (1999). National and international geospatial data policies, in *Geographical Information Systems: Principles, Techniques, Applications and Management*, P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, eds., Wiley, New York, pp. 767-787.

Star, S. L., and K. Ruhleder (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces, *Information Systems Research* 7(1), 111-134.

Tosta, N., and M. Domaratz (1997). The U.S. National Spatial Data Infrastructure. in *Geographic Information Research: Bridging the Atlantic,* M. Craglia and H. Couclelis, editors, Taylor and Francis, London, pp. 19-27.

Tsou, M., and B. P. Buttenfield (1998). Client/server components and metadata objects for distributed geographic information services, *Proceedings, GIS/LIS '98, Fort Worth, TX, November.*

White, D., A. J. Kimerling, and W. S. Overton (1992). Cartographic and geometric components of a global sampling design for environmental monitoring, *Cartography and Geographic Information Systems 19*(1), 5-22.

Worboys, M. F. (1995). *GIS: A Computing Perspective*, Taylor & Francis, London.

APPENDIXES

APPENDIX A
WORKSHOP PARTICIPANTS

Prudence S. Adler, Assistant Executive Director, Association of Research Libraries

Timothy M. Alexander, Consultant, Washington, D.C.

William Y. Arms, Vice President, Corporation for National Research Initiatives

James Barrett, Space Imaging Corporation

Gerald S. Barton, National Oceanic and Atmospheric Administration

Francis P. Bretherton, Director, Space Science and Engineering Center, University of Wisconsin

Don Buhler, Bureau of Land Management

Barbara P. Buttenfield, Department of Geography, University of. Colorado

Nicholas Chrisman, Department of Geography, University of Washington

David J. Coleman, Chairman, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Canada

Donald F. Cooke, President, Geographic Data Technology, Inc.

Frederic W. Corle II, Partner, Potomac Research Group

Max J. Egenhofer, Department of Spatial Information, Science and Engineering, University of Maine

John D. Evans, Department of Urban Studies and Planning, Massachusetts Institute of Technology

Christos Faloutsos, Computer Science Department, Carnegie Mellon University

Gary Fitzpatrick, Library of Congress

Michael J. Folk, National Center for Supercomputing Applications, University of Illinois

Kenneth E. Foote, Department of Geography, University of Texas

Alan M. Gaines, Earth Sciences Division, National Science Foundation
Kenn Gardels, Director, Research Program in Environmental Planning and Geographic Information Systems, University of California, Berkeley
Henry L. Garie, Department of Environmental Protection, State of New Jersey (President, National States Geographic Information Council)
Barry J. Glick, President, GeoSystems Global Corporation (MapQuest)
Michael F. Goodchild, National Center for Geographic Information and Analysis, University of California, Santa Barbara
Stephen Griffin, Digital Libraries Initiative, National Science Foundation
Wayne A. Hallada, Defense Intelligence Agency
Linda L. Hill, Alexandria Digital Library, University of California, Santa Barbara
Thomas M. Holm, EROS Data Center, U.S. Geological Survey
Sally Howe, Associate Director, National Coordination Office for Computing, Information, and Communications
Robert E. Kahn, President, Corporation for National Research Initiatives
Thomas A. Kalil, Senior Director, National Economic Council, Executive Office of the President
Karen Kemp, National Center for Geographic Information and Analysis, University of California, Santa Barbara
Annette J. Krygiel, Institute for National Strategic Studies, National Defense University
Nand Lal, Digital Libraries Initiative, NASA Goddard Space Flight Center
Nina S-N. Lam, Department of Geography, Louisiana State University
Roberta E. Lenczowski, Deputy Director for Operations, National Imagery and Mapping Agency
Michael E. Lesk, Director, Division of Information and Intelligent Systems, National Science Foundation
Clifford A. Lynch, Executive Director, Coalition for Networked Information
David M. Mark, Department of Geography, State University of New York, Buffalo
Patrick McGlamery, Map and Geographic Information Center, University of Connecticut
Eric Miller, Online Computer Library Center, Inc.
John Moeller, Executive Director, Federal Geographic Data Committee, U.S. Geological Survey
Joel Morrison, Geography Division, Bureau of the Census

Douglas D. Nebert, Federal Geographic Data Committee, U.S. Geological Survey

Robert Neches, Information Sciences Institute, University of Southern California

Harlan J. Onsrud, Department of Spatial Information Science and Engineering University of Maine

Hedy C. Rossmeissl, National Mapping Division, U.S. Geological Survey

Walter Senus, Chief Scientist, National Imagery and Mapping Agency

Ben A. Shneiderman, Department of Computer Science, University of Maryland

Karen C. Siderelis, Director, North Carolina Center for Geographic Information and Analysis

Terence Smith, Alexandria Digital Library, University of California, Santa Barbara

Gale TeSelle, Information Technology Division, Natural Resources Conservation Service

Rex W. Tracy, Vice President, GDE Systems, Inc.

Robert Tufts, TASC

A. Keith Turner, Department of Geology and Geological Engineering, Colorado School of Mines

Ferris Webster, College of Marine Studies, University of Delaware (Chair, ICSU Panel on World Data Centers)

Frederick Weingarten, Computing Research Association

James F. Williams II, Dean of Libraries, University of Colorado

Maria Zemankova, Division of Information and Intelligent Systems, National Science Foundation

NRC Staff

Thomas M. Usselman, Study Director, Mapping Science Committee, National Research Council

Jennifer T. Estep, Administrative Assistant, Mapping Science Committee, National Research Council

Molly Vogt, Workshop Assistant, National Geographic Society

APPENDIX B
Contributed White Papers

The following white papers were prepared and distributed to the workshop participants as provocative authored articles; they were designed to stimulate thought and were not peer reviewed. Authors did not have the opportunity to make revisions after the workshop. The papers (generally two to three pages long) are available on the World Wide Web at *http://www4.nas.edu/cger/besr.nsf* following hyperlinks to Mapping Science Committee, reports, distributed geolibraries, white papers. This web site will be maintained at least through the end of 2000.

What Is a Geolibrary? Michael F. Goodchild, University of California, Santa Barbara

What Is a Geolibrary? Clifford A. Kottman, Open GIS Consortium

Making the Case for Distributed GeoLibraries Barbara P. Buttenfield, University of Colorado, Boulder

Confidence in Distributed Digital Geolibraries Roberta E. Lenczowski, National Imagery and Mapping Agency

Data Quality in Distributed GeoLibraries Rex W. Tracy, GDE Systems, Inc.

National Satellite Land Remote Sensing Data Archive: "A National Asset" Thomas M. Holm, U.S. Geological Survey EROS Data Center

Possible Research Topics Related to Geolibraries Michael F. Goodchild, University of California, Santa Barbara

Distributed Geolibraries: Challenges and Opportunities from a Computing and Scientific Data Management Perspective Mike Folk, National Center for Supercomputer Applications, University of Illinois

Geolibrary and Statewide Electronic Atlas, Nina Lam, Louisiana State University

Putting The User First: Implications for the Geolibrary A. Keith Turner, Colorado School of Mines

Geolibraries Walter Senus, National Imagery and Mapping Agency

Producer Rights and Public Rights in Spatial Data Sets in Geolibraries Harlan J. Onsrud, University of Maine

Distributed Geolibraries Patrick McGlamery, University of Connecticut

Geolibrary Access of Historical Climate Data Publications in the NOAA Library Gerald S. Barton, National Oceanic and Atmospheric Administration

Geolibraries: Integration into to the Life Cycle of Information Creation and Use Linda L. Hill, Alexandria Digital Library Project, University of California, Santa Barbara

The Geodata Network Kenn Gardels, University of California, Berkeley

What is a Geolibrary? Christos Faloutsos, Carnegie Mellon University

APPENDIX C
Workshop Agenda
Workshop on Distributed Geolibraries: Spatial Information Resources

Mapping Science Committee
National Research Council - National Academy of Sciences
Room Green 130, 2001 Wisconsin Avenue, NW
Washington, D.C.

The workshop is intended to address the following:

- Development of a *vision* for geospatial data dissemination and access in 2010.
- Comparison of *current* efforts in digital library research, clearinghouse development, and other data distribution and search activities.
- Suggestion of short-term and long-term *research and development* needed to achieve the vision.
- Identification of the **policy and institutional issues**, particularly for convergence of efforts to realize the vision.

Monday, June 15, 1998

8:00 Registration, Coffee, Continental Breakfast

Plenary

8:30 INTRODUCTION AND PURPOSE
Mike Goodchild, Workshop Chair, University of California, Santa Barbara

9:00 FRAME OF REFERENCE—Digital Libraries, Internet, Information Sciences

- **Robert Kahn**, President, Corporation for National Research Initiatives
- **Michael Lesk**, Division Director, Information and Intelligent Systems, National Science Foundation

10:20 BREAK

10:40 POLICY-INSTITUTIONAL-EDUCATIONAL

- **Eric Miller**, Online Computer Library Center, Inc.
- **Harlan J. Onsrud,** Department of Spatial Information Science and Engineering, University of Maine

12:00 LUNCH

1:00 **Breakout Sessions**

Each group will discuss the following questions:

1. What is a suitable vision for geospatial data dissemination and access in 2010 (code named geolibraries)?
   - What is a geolibrary?
   - What types of information might a geolibrary contain?
   - What services might it offer?
   - What types of users would there be?
   - What kinds of access restrictions might be needed?
   - Should a geolibrary be integrated with other information services?
2. Policy and institutional issues
   - What are the legal, ethical, and political issues involved in creating geolibraries? For example, what problems could geolibraries raise related to intellectual property rights? How might such issues affect the technical development of geolibraries?
   - Who should pay for the creation and maintenance of geolibraries? What components might be "free" (funded by the public sector or by the private sector as loss leaders)? What institutional structures would be needed for geolibraries? What organizations might take a lead in their development?

- What are the cognitive problems associated with using geolibraries? Is it possible to construct a geolibrary that is useful to a child in grade 3, for example? What protocols would users need to master, and what problems would occur in using geolibraries across cultural or linguistic barriers? What are the implications of a national-level distributed geolibrary on education? What are the prospects for international geolibraries?

4:00 **Plenary**
Rapporteurs will present results of each breakout group.

5:00 ADJOURN
5:30 RECEPTION; Followed by Dinner at 6:30

Tuesday, June 16, 1998

8:00 Coffee, Continental Breakfast

Plenary

8:30 CURRENT ACTIVITIES AND TECHNICAL ISSUES

- **Ben Shneiderman**, Department of Computer Science, University of Maryland
- **Thomas Kalil**, Senior Director, National Economic Council, The White House
- **Terence Smith,** Director, Alexandria Project, University of California, Santa Barbara

10:20 BREAK

10:30 **Breakouts**
Each group will discuss the following questions:

3. Ongoing Activities
   - What components of existing efforts in digital library research, clearinghouse development, and other data distribution

and search activities might form a part of a distributed geolibrary system? KDI? Digital Earth?

- Do the necessary data sets to support geolibraries exist? What initiatives are needed to develop or compile them?
- How could the geolibraries concept be expanded beyond the national level to take advantage of international and global information resources?

4. Technical Issues and R&D Needs
   - Integration of geospatial data across themes and scales.
   - A new generation of search engines
   - Geospatial interoperability
   - User interface metaphors
   - Collection-level metadata
   - Which of the R&D needs can be attained in the next few years, and which ones may take 5 to 10 years?

12:00 LUNCH

1:30 **Plenary**

- Results of the research needs and current activities to realize a "geolibrary" vision
- Overall Workshop Results

4:30 **Adjourn**

Appendix D
Example Prototypes

Detailed in this section is a sample of World Wide Web sites chosen to illustrate existing elements of the distributed geolibrary vision. Each is largely isolated from one another and falls short of the full vision. Taken together, the set illustrates both what is already possible and how far we still are from a distributed geolibrary.

Microsoft's Terraserver (*www.terraserver.com*)

Terraserver offers digital imagery (Figure D.1) from the Russian SPIN-2 satellites and digital orthorectified photographs (orthophotos) from the U.S. Geological Survey. The archive contains over 1 terabyte of information and can be queried by pointing, zooming, and panning on a basemap or by specifying place-names. No services are provided for finding or integrating other data based on place.

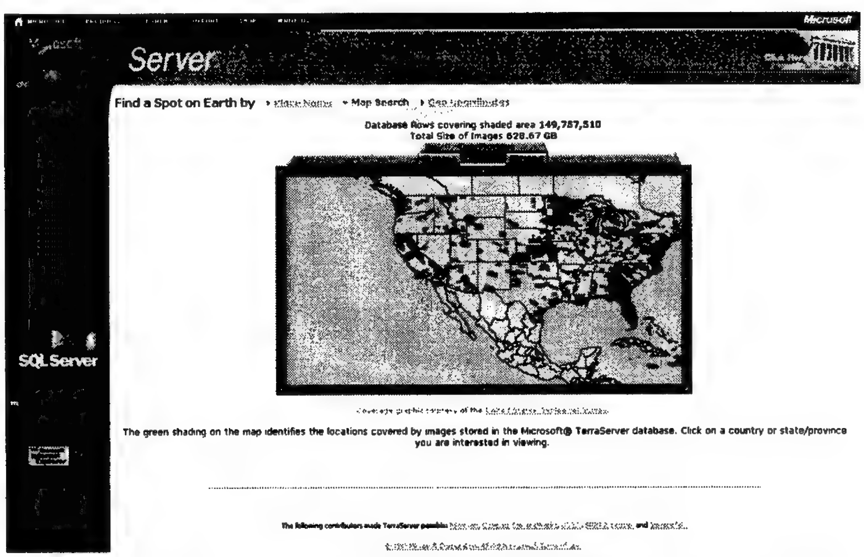


FIGURE D.1. U.S. coverage in the terraserver database in October 1998.

MapQuest (*www.mapquest.com*)

The MapQuest site offers a range of services based on its database (see Figure D.2). The map illustrates the ability to provide services based on specific collections, in addition to serving unmodified information.

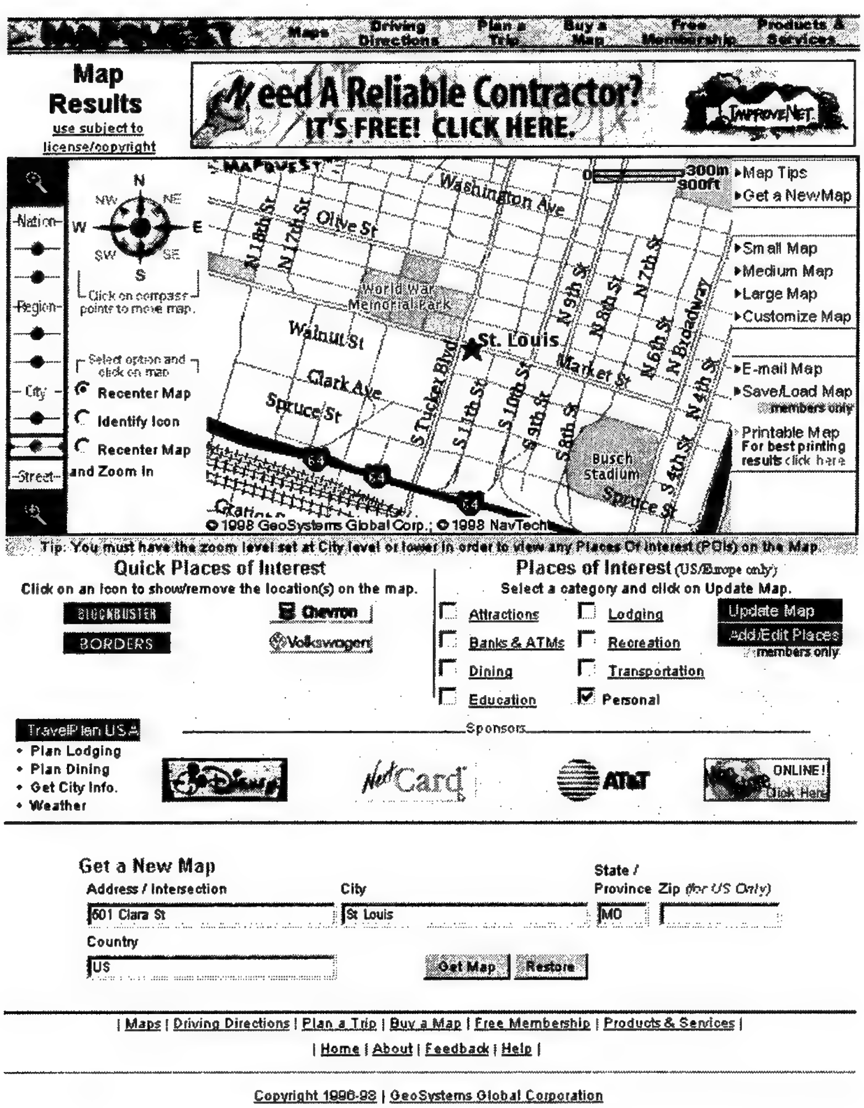


FIGURE D.2. Result of a request to the MapQuest site for a map centered on an address in St. Louis, Missouri.

Environmental Protection Agency ZIP Code Search

(*www.epa.gov/enviro/zipcode_js.html*)

The U.S. Environmental Protection Agency's (EPA) web site offers several forms of place-based search through the agency's archives, including Maps On Demand and ZIP code search. From the web site:

> The EPA Envirofacts Warehouse is a database that includes information on Superfund sites, drinking water, air pollution, toxic releases, hazardous waste, and water discharge permits. Through Envirofacts, you can get lists of which facilities in your neighborhood are releasing pollutants or are legally handling hazardous materials, where any Superfund sites are located and what their cleanup status is, and more. In many cases, you can link to more information about the chemicals involved at the listed sites, and find out whether they are potentially harmful.
>
> Through Envirofacts' EnviroMapper feature, you can customize a computer-generated map of your neighborhood to view the location of EPA regulated sites, schools, churches, streams, streets, and other geographic features.

U.S. Bureau of the Census TIGER Map Server

(*tiger.census.gov*)

The Bureau of the Census web site supports place-based search for census data. In Figure D.3 the TIGER Map Server has generated a map of part of Goleta, California, showing features selected by the user. The main purpose of the TIGER Map Service project is to provide a good-quality, national scale, street-level map to users of the World Wide Web. This service is freely accessible to the public, and based on an open architecture that allows other Web developers and publishers to use public domain maps generated by this service in their own applications and documents. We planned to provide high-quality street maps, with simple GIS capabilities such as point display and statistical choropleth mapping.

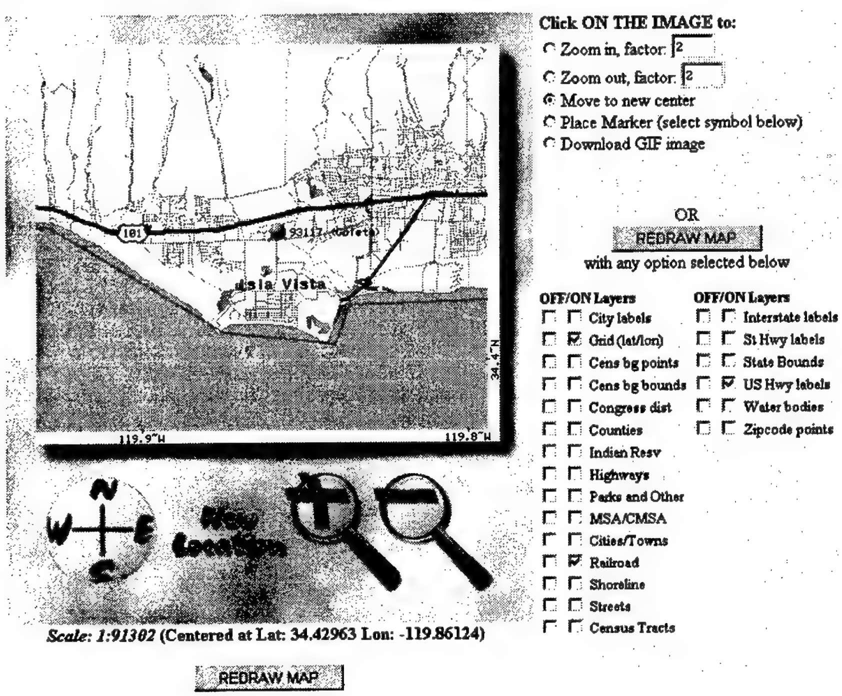


FIGURE D.3. The U.S. Bureau of the Census TIGER Map Server will create maps to user specifications, based on place-based search.

U.S. Geological Survey National Atlas

(*www.usgs.gov/atlas*)

The National Atlas web site creates and delivers maps on demand from the National Atlas database. Figure D.4 shows the initial stage of specification.

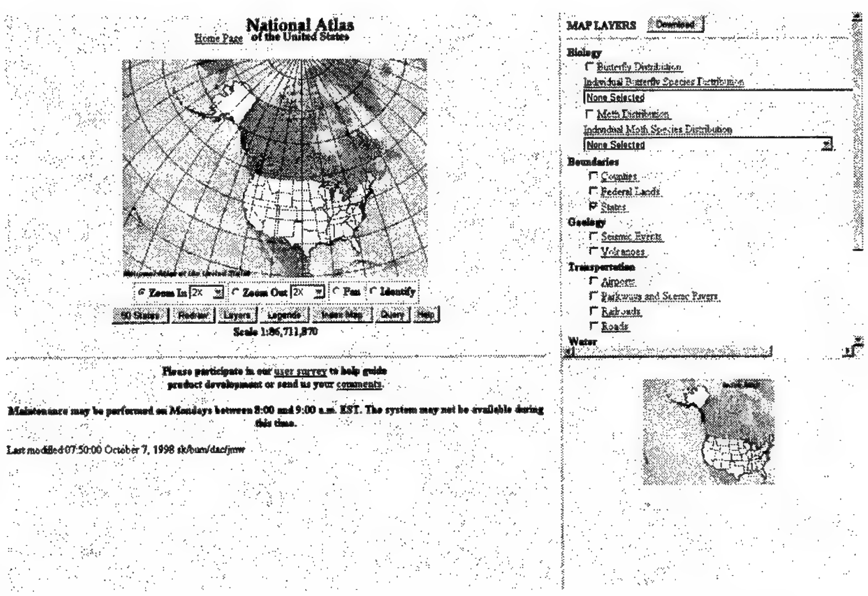


FIGURE D.4. The U.S. Geological Survey's National Atlas web site, which allows users to specify and create customized maps from the National Atlas database.

MIT's Digital Orthophoto Server (*ortho.mit.edu*)

This Massachusetts Institute of Technology web site serves digital orthophotos (DOQs) for the area around Boston. Figure D.5 shows the index page; by clicking on a tile it is possible to retrieve the associated data at a user-defined level of resolution.

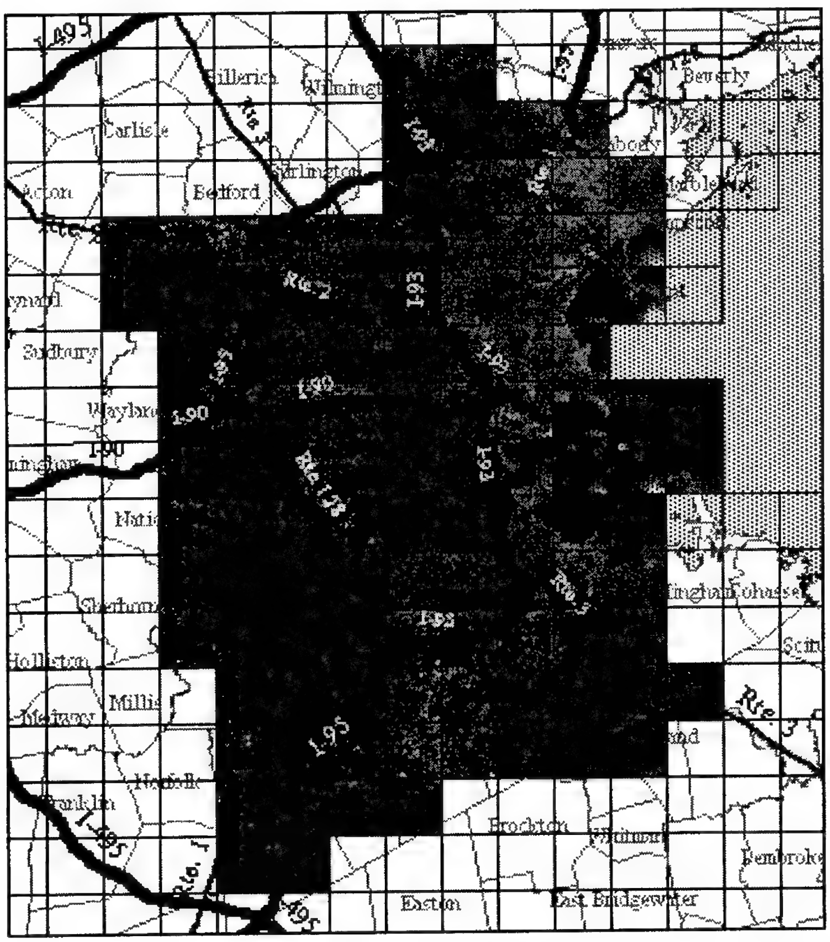


FIGURE D.5. Index page for MIT's Digital Orthophoto server, which provides downloadable orthophotography for the Boston area.

Alexandria Digital Library (*alexandria.ucsb.edu*)

The Alexandria Digital Library (ADL) is the product of a research project at the University of California, Santa Barbara, funded through the Digital Library Initiative of the National Science Foundation, the National Aeronautics and Space Administration, and the Defense Advanced Research Projects Agency. Figure D.6 shows the first screen in ADL's process of defining a place-based search. Additional properties can be specified to narrow the search, which is then applied to the order 10^6 data sets in the current ADL collection.

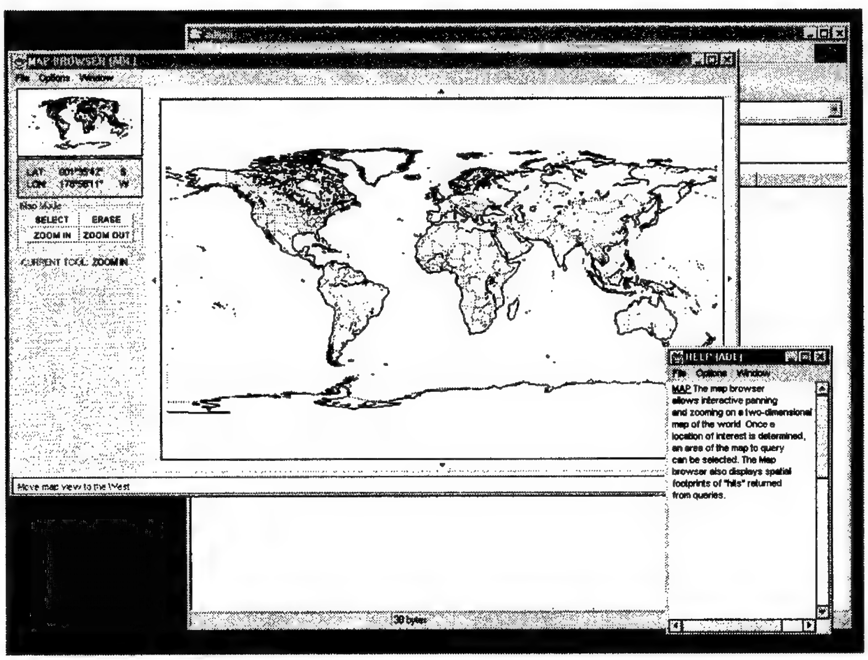


FIGURE D.6. Opening screen of the Alexandria Digital Library's Java-based engine for searching its collection based on location.

Microsoft's HomeAdvisor (*www.homeadvisor.com*)

This Microsoft web site is designed to help people searching for homes. It includes the ability to access demographic and other information about neighborhoods (see Chapter 1), search listings, and estimate payments. Figure D.7 shows the opening screen.

FIGURE D.7. Microsoft's HomeAdvisor web site.

APPENDIX E
Biographical Sketches of Panel Members

Michael F. Goodchild (*Chair*) is professor and chair of the Department of Geography at the University of California, Santa Barbara; director of the National Center for Geographic Information and Analysis; and associate director of the Alexandria Digital Library. He received his B.A. in physics from Cambridge University and Ph.D. in geography from McMaster University. Dr. Goodchild taught at the University of Western Ontario for 19 years before moving to his present position in 1988. His research interests focus on the generic issues of geographic information, including accuracy and the modeling of uncertainty, design of spatial decision support systems, development of methods of spatial analysis, and data structures for global geographic information systems. His publications include the two volume text entitled *Geographical Information Systems: Principles, Techniques, Applications and Management* (1999, Wiley). He is also chair of the Mapping Science Committee.

Prudence S. Adler is assistant executive director of the Association of Research Libraries (Washington, D.C.), where she is primarily responsible for federal relations and information policy activities. Much of her recent emphasis has been on intellectual property rights in an electronic environment.

Barbara P. Buttenfield is associate professor of geography at the University of Colorado in Boulder. She holds a Ph.D. in geography from the University of Washington. She has served on the faculty at the State University of New York, Buffalo; the University of California, Santa Barbara; and the University of Wisconsin, Madison. Dr. Buttenfield's research interests focus on cartographic knowledge construction, spatial data delivery on the Internet, and visualization tools for geographic modeling. A current project to evaluate user

interface tools for the Alexandria Digital Library is funded jointly by NSF, ARPA, and NASA. She is past President of the American Cartographic Association, and serves on the editorial boards of *Computers Environment and Urban Systems*, *Transactions on GIS*, and *Cartographic Perspectives*. She is also a member of the Mapping Science Committee.

Robert E. Kahn is chairman, CEO, and president of the Corporation for National Research Initiatives (CNRI), which he founded in 1986 after a 13 years at the U.S. Defense Advanced Research Projects Agency (DARPA). CNRI provides leadership and funding for research and development of the National Information Infrastructure. Dr. Kahn earned a Ph.D. degree from Princeton in 1964. He worked at Bell Laboratories and as an assistant professor of electrical engineering at MIT. He took a leave of absence from MIT to join Bolt Beranek and Newman, where he was responsible for the system design of the Arpanet. In 1972 he moved to DARPA and subsequently became director of DARPA's Information Processing Techniques Office (IPTO). While director of IPTO he initiated the United States government's billion-dollar Strategic Computing Program, the largest computer research and development program ever undertaken. Dr. Kahn conceived the idea of open-architecture networking. He is a coinventor of the TCP/IP protocols and was responsible for originating DARPA's Internet program. Dr. Kahn also coined the term *national information infrastructure* (NII) in the mid-1980s, which later became more widely known as the information superhighway. His recent work has been developing the concept of a digital object infrastructure to provide a framework for interoperability of heterogeneous information systems, particularly as applied to digital libraries. Dr. Kahn is a member of the National Academy of Engineering and a 1997 recipient of the National Medal of Technology.

Annette J. Krygiel is with the Institute for National Strategic Studies, National Defense University, Ft. Lesley J. McNair, in Washington, D.C. Dr. Krygiel has a B.S. in mathematics from St. Louis University, and a M.S. and Ph.D. in computer science from

Washington University, St. Louis. In her doctoral research she developed modeling techniques for parallel computing architectures. She began her government career in 1963, serving with the Defense Mapping Agency (DMA) until July 1994. While at DMA her areas of endeavor included software development, software engineering, management of research initiatives in computer science and telecommunications, and program management of large-scale systems. Dr. Krygiel rejoined DMA's special program office to manage the program integration, test and delivery phases of DMA's Digital Production System, one of the U.S. Department of Defense's (DOD) largest software developments. Subsequently, Dr. Krygiel served as DMA's chief scientist until her formal appointment by the Secretary of Defense as the Director of the Central Imagery Office (CIO), a DOD combat support agency. She remained as Director for twenty-seven months until that agency merged into the National Imagery and Mapping Agency in October 1996. She was awarded the National Intelligence Distinguished Service Medal while CIO Director. Dr. Krygiel was subsequently appointed to the Institute for National Strategic Studies at the National Defense University, where she is investigating the problem of large-scale system integration.

Harlan J. Onsrud is associate professor in the Department of Spatial Information Science and Engineering at the University of Maine and chair of the Scientific Policy Committee of the National Center for Geographic Information and Analysis. He received B.S. and M.S. degrees in Civil Engineering from the University of Wisconsin and a Juris Doctorate from the University of Wisconsin Law School. His research focuses on (1) analysis of legal and institutional issues affecting the creation and use of digital databases and the sharing of geographic information, (2) assessing utilization of GIS and the social impacts of the technology, and (3) developing and assessing strategies for supporting the diffusion of geographic information innovations. He is also a member of the Mapping Science Committee.